HAUTE AUTORITÉ DE SANTÉ

SUMMARY REPORT

Indirect comparisons
Methods and validity

July 2009

This executive summary can be downloaded from
www.has-sante.fr

Haute Autorité de Santé
Service communication
2 avenue du Stade de France - F 93218 Saint-Denis La Plaine CEDEX France
Tel. :+33 (0)1 55 93 70 00 -Fax:+33 (0)1 55 93 74 00

Contents

Summary 6

Definition 6
Context and potential benefits 6
Statistical methods 6
Validity of indirect comparisons 7
Are direct comparison trials the gold standard? 7
Current use and acceptance by agencies 7
Critical review 7
Conclusions - Key points 7

1. Introduction 9

1.1 Definitions and general comments 9
1.2 Objectives 10

2. Background 11

2.1 Lack of direct comparisons between active therapies 11
2.2 Predilection for placebo-controlled trials 12
2.3 Nontransitivity of statistical tests 12
2.4 Non-inferiority clinical trials 13

3. Literature review methods and selection criteria 14

3.1 Information sources 14
3.2 Literature search strategy 14
3.3 Publication selection criteria 15

4. Various methods used to carry out indirect comparisons 16

4.1 Inappropriate methods 16
4.2 Adjusted indirect comparisons 18
4.3 Mixed approach for direct and indirect comparisons 22
4.4 Methods based on a global estimate from a network of trials 24
4.5 Bayesian network meta-analysis 27
4.6 Estimation using a mixed linear model 30
4.7 Meta-regression 33
4.8 Methods for continuous endpoints 35
4.9 Comparison of the different methods 35

5. Current use of indirect comparisons 37

5.1 Bibliographical data 37
5.2 Some recent publications 37
5.3 Use in health technology assessments and guidelines 39

6. Validity of indirect comparisons 40

6.1 How valid are indirect comparisons? 40
6.2 Validity studies 40
6.3 Is direct comparison still the gold standard? 42
6.4 Outlook 44

7. Position of regulatory bodies 45
8. Interpretation of results 46
9. Critical review guide 47
9.1 Ideal situation 47
9.2 Critical review guide 47
9.3 Case studies 50
10. Conclusion 55
11. Key points 56
Annex 1. List of participants 57
Annex 2. References 58
Annex 3. Search for guidelines, health technology assessments and methodological procedures on the Internet; list of websites consulted 64
Index 65

Abbreviations

MA	Marketing authorisation
HAS	*Haute Autorité de Santé*
FDA	Federal Drug Administration
EMA	European Medicines Agency
AFSSAPS	*Agence Française de Sécurité Sanitaire des Produits de Santé*
NICE	National Institute for Health and Clinical Excellence

Summary

Definition

Indirect comparison methods are used to measure the effect of treatment A compared with treatment B based on the results of trials of A and of B versus the same control (placebo or active treatment). So these methods provide extrapolations based on the hypothesis that the effects that A and B would have *(*compared with a control in a head-to-head trial) are the same as those observed in trials used in the indirect comparison.

Direct comparisons compare A and B directly in a head-to-head trial.

The results of direct comparisons can be combined with those of indirect comparisons by using a mixed approach, known as a mixed treatment comparison (MTC).

Context and potential benefits

Indirect comparisons are being used more and more often because they make it possible to address common issues in evaluation concerning the hierarchy of efficacy and/or safety of competing treatments.

In many clinical fields, competing treatments are assessed against placebo and direct comparisons are rare. Indirect comparisons can make it possible to estimate the relative efficacy and/or safety of therapies in relation to each other before any direct comparison trials are available.

In some cases, direct comparisons may be impossible for ethical reasons (e.g. in life-threatening disorders or those causing short-term disability) or for feasibility reasons in the case of rare diseases. In such clinical fields, indirect comparison is the only method that can be used to estimate the relative efficacy and/or safety of treatments.

In addition, the description of direct comparison trials as the "gold standard" is debatable, as their methodological quality is often lower than that of placebo-controlled trials (inadequate blinding, lack of sensitivity, etc.).
A direct comparison may also reveal problems related to the non-transitivity of certain statistical tests in two-by-two comparisons.

Statistical methods

Appropriate indirect comparison methods have been developed over the last few years. Many studies intended to establish an indirect comparison still use unsuitable methods such as naive comparisons of point estimates or of active arms of different controlled studies. These methods are not adequate and studies using them should not be called indirect comparisons.

There are two types of appropriate indirect comparison method:

- methods which only compare two treatments simultaneously, such as adjusted indirect comparisons. To compare multiple treatments, multiple two-by-two comparisons are needed;
- methods which include a number of treatments simultaneously using a complex statistical model. They model the network of possible comparisons between all competing treatments. These methods include trials comparing two treatments and multi-arm trials. Different methods for estimating the parameters of these models have been proposed, such as Bayesian methods, mixed linear models and meta-regression.

There has been little research on the relative performances of these different methods. A simulation study has shown that all these methods appear free of bias.

In the current state of knowledge, a Bayesian network meta-analysis seems to be the most useful method as it is flexible and attractive, at least at the theoretical level.

To avoid arbitrary selection of data for analysis, whatever method is used, calculations of indirect comparisons should be applied to the results of meta-analyses, which assumes an in-depth data search and selection of trials based on their ability to produce an unbiased estimate.

Validity of indirect comparisons

There is only one wide-ranging study of empirical comparisons of results derived from indirect comparisons with those derived from direct comparisons[1]. This validity study concerns a method for adjusted indirect comparisons. The criterion for discrepancy used was the presence of a statistically significant difference between estimates of the treatment effect of the two approaches. Of the forty-four cases studied (use of two treatments for a given disorder), only three cases of discrepancy were found. Differences were in both directions. No factors predictive of discrepancy were identified.
The other validity studies available address only one clinical field, and are therefore more valuable as case studies than as empirical validation studies.
Finally, first available data tend to conclude that indirect comparisons are valid, but further empirical research is needed to reach a definite conclusion on this point.

Are direct comparison trials the gold standard?

Empirical validation studies of indirect comparisons start from the currently accepted premise that head-to-head studies are the gold standard and that the estimates they provide are free from bias. However, this premise has not been validated by any empirical study. This is important, as direct comparison trials of two active treatments are often carried out as open trials, and there have been cases where results could not be replicated depending on the trial sponsor.

Current use and acceptance by agencies

The use of indirect comparison methods is no longer unusual. More than a hundred publications (scientific articles, guidelines, health technology assessments) using this approach are now available, contributing substantial empirical experience. However, there is still little critical assessment.
At registration level, none of the agencies (FDA, EMA, AFSSAPS) has adopted a formal position on the acceptability of justifications based on the results of indirect comparisons.
Of the organisations that produce health technology assessments or guidelines, only NICE mentions in its procedures the possibility of using indirect comparison, as an experimental approach and on a case by case basis. It is explicitly stated that the standard approach remains direct comparison.

Critical review

As a result of theoretical study of the methods proposed and experience acquired with the first uses of these methods, it is possible to produce a list of key points to be analysed in a critical review of the results of indirect comparisons. A working critical review guide is proposed at the end of the literature study.

Conclusions - Key points

To make up for the lack of comparative trials between active treatments in a very wide range of clinical fields, the use of indirect comparisons may be considered whenever there is a question of a hierarchy of efficacy and/or safety of competing treatments.

This approach may make it possible to assess the position of a new treatment in relation to existing treatments as soon as it has been licensed, without having to wait several years after the treatment has been introduced for more conventional comparative data from head-to-head trials.

[1] Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 2003;326(7387):472.

Results from direct comparisons and those from indirect comparison should be carefully compared and if possible combined using a mixed treatment comparison.

This approach should be implemented using network meta-analyses, which are the best way of establishing a complete hierarchy for a given clinical field (including all available treatment resources, if possible). Other valid methods may also be used while waiting for in-depth comparisons of the relative performance of the different methods.

As there is currently only one large study of the validity of indirect comparisons, further empirical verifications should be carried out using modern indirect comparison methods. Research efforts should be continued to study issues related to the reliability of direct comparisons and the performance of the different indirect comparison methods proposed.

More widespread use of indirect comparison methods could lead to better strategies for establishing a hierarchy of treatments in relation to each other, based on factual data.

Consequences for practice

- Indirect comparisons may be carried out as the foundations for a critical evaluation of their quality exist.
- The use of indirect comparisons may be considered whenever there is a question of a hierarchy of efficacy or safety of competing treatments.
- The acceptability of indirect comparisons needs to be considered on a case by case basis, taking account of the clinical field and the consequences of a lack of direct comparison trials.

Consequences for research

- Research to compare the performance of the different statistical methods is needed to establish the preferred method(s).
- Empirical research to validate direct comparison is also needed to obtain a more accurate picture of the appropriateness of regarding it as the gold standard for comparing two active treatments.
- Further empirical validation studies of networked indirect comparisons are also needed.

1. Introduction

1.1 Definitions and general comments

Direct and indirect comparisons are two approaches used to compare the efficacy (or the safety) of two active treatments, A and B, in order to decide whether A is superior or non-inferior to B.

A direct comparison is carried out when the two treatments A and B are compared directly in a randomised clinical trial. This type of trial is also called an active-control trial, or a head-to-head comparison. The result of this type of trial is a measure of the effect of A compared with B (relative risk, odds ratio, risk difference, hazard ratio, absolute or relative difference in means, etc.) and a statistical test of the statistical reality of a difference of effect between A and B. In this way, a trial directly comparing a new treatment to an accepted standard treatment demonstrates the superiority or the non-inferiority (depending on the trial objective) of the new treatment in relation to the standard.

For indirect comparison, the efficacy of the two treatments A and B is compared by the intermediary of the respective efficacy of both treatments compared with a common control, usually placebo[2]. The idea is to decide whether A is superior or inferior to B based on the result of comparing A with the control and the result of comparing B with the same control (Figure 1). This comparison of A and B is indirect as it makes use of a third treatment, the common control. Trials of A versus control and of B versus control provide the basic comparisons from which an indirect comparison of A with B can be made.

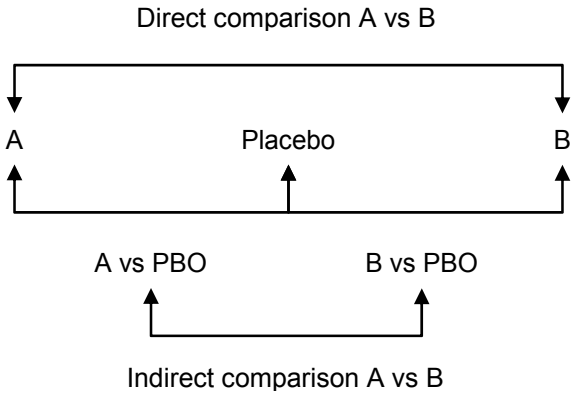


Figure 1 - Illustration of the principles of direct and indirect comparisons. In this example, the control treatment for the basic comparisons is placebo (PBO)

As an example, when the endpoint is binary, results from a direct comparison may be expressed in the form of relative risk This relative risk measures the additional reduction in frequency of the event, the endpoint, contributed by treatment A compared with that contributed by treatment B (B being effective and already reducing the frequency of the event).
Table 1 illustrates the typical results of a direct comparison of A with B.

Table 1 - Results of a direct comparison trial of A versus B

	Treatment A n=1 000	Treatment B n=1 000	Relative risk (95% CI)	p-value
Deaths	127 (12.7%)	234 (23.4%)	0.54 (0.45-0.66)	< 0.001

[2] Indirect comparisons may also be carried out with a common active control.

The results of the direct comparisons shown in Table 2 give us two relative risks, i.e. A versus PBO and B versus PBO. These relative risks measure the absolute efficacy of the treatment in question (A or B), i.e. the relative reduction in frequency of the event obtained by using this treatment compared with not using it[3].
Intuitively, if the relative risk of A versus PBO is smaller than that of B versus PBO, there is a tendency to consider that A is more effective than B, although, as we will see later on, this method of comparison is inappropriate.

Table 2 - Results of basic direct comparison trials required for an indirect comparison of A and of B via the intermediary of placebo

A versus Placebo

	Treatment A n=1 000	Placebo n=1 000	Relative risk (95% CI)	p-value
Deaths	134 (13.4%)	274 (27.4%)	0.49 (0.41-0.59)	< 0.001

B versus Placebo

	Treatment B n=1 000	Placebo n=1 000	Relative risk (95% CI)	p-value
Deaths	56 (5.6%)	103 (10.3%)	0.50 (0.4-0.74)	< 0.001

Until recently, direct comparison was regarded as the reference approach, while there were many reservations about indirect comparisons, particularly in view of the simplistic methods used. Appropriate methods have been developed. Theoretical and empirical validation studies in this area carried out over the last few years suggest that there is a role in evaluation for properly conducted and rigorously interpreted indirect comparisons.

1.2 Objectives

This literature review has a number of objectives:

1. present the background to the development of indirect comparison methods;
2. identify situations in which indirect comparisons can make a real contribution;
3. present the various reliable indirect comparison methods;
4. describe the main advantages and limitations of each method;
5. examine the validity of indirect comparisons;
6. propose an evaluation and critical review method for indirect comparison studies.

[3] In the most common case where the endpoint has the value of treatment failure, RR < 1 indicates that the treatment tested is superior to the control treatment. The closer the relative risk is to zero, the greater the benefit contributed by the test treatment. RR > 1 indicates a harmful effect of the test treatment or superior efficacy of the control treatment. The interpretation is reversed if the endpoint is a favourable event. The relative risk reduction (RRR) is quite frequently used instead of the relative risk : RRR = (1-RR)x100%.

2. Background

Number of competing therapies exist for the same disease or disorder. These are the therapies (drugs or other types of therapy) intended for the same patients, offering a number of treatment options. For example, during the acute phase of myocardial infarction, a number of thrombolytics may be used to reopen the obstructed coronary artery; primary angioplasty has the same objective. An acute migraine attack may be treated with non-specific therapies (analgesics, non-steroidal anti-inflammatory drugs) or with specific drugs such as triptans. A number of anticholinesterase agents are indicated in Alzheimer's disease.

The aim of evaluation is to compare competing therapies with each other to assess their relative efficacy and safety. In order to make rational choices or recommendations, it is useful to know whether one treatment is more effective and/or better tolerated than the others and if so, how much more effective and/or safe it is compared with the others. The idea is to be able to offer the patient the best first-line treatment and to reserve the others for second-or third-line treatment.

Unfortunately, direct comparisons of competing therapies are often not available, making it impossible to carry out an in-depth comparison of the relative benefits and risks of the various therapies. The lack of this information is an obstacle to the rational process of evidence-based evaluation of therapies. Therapies are then chosen on the basis of their pharmacological properties, the results for intermediate endpoints, and various speculative considerations, and often it is external factors (attraction of novelty, marketing pressure, etc.) which are used to select the preferred therapy.
This was the situation in which the concept of indirect comparison began to appear. Indirect comparison involves making the best use of information obtained from available direct comparisons (which are mainly derived from placebo-controlled trials, possibly from a few rare trials versus active therapies) to estimate the relative efficacy and safety of the different competing therapies (1-4) .

2.1 Lack of direct comparisons between active therapies

What are the reasons that make many direct comparison trials between active therapies needed for evaluation unavailable?

In an ideal world this question would not have to be asked. Once a first therapy has been shown effective, any new treatments would have to be compared to it to demonstrate their superiority (in terms of efficacy or safety). So in this ideal situation, it would be possible to establish a hierarchy of all therapies: the first therapy T1 demonstrates its efficacy compared with placebo (unless this is impossible for ethical reasons). It becomes the standard treatment. The second therapy T2 is compared to T1 and demonstrates its superiority (in terms of efficacy and/or safety). It becomes the new standard treatment. A third therapy is compared with T2, etc.

However, even if this process is fully complied with, direct comparisons (absent in this case) would nevertheless have been useful. This happens for example when the third treatment fails to demonstrate its superiority over T2. What happens to T3? Should it be abandoned completely? Couldn’t it become an alternative second-line therapy (e.g. for patients who cannot tolerate T2)?. It might actually be more effective and/or better tolerated than T1 (although not superior to T2). A trial of T3 versus T1 would be needed to provide the answer. This would imply treating patients with T1, a therapy already known not to be optimal at the time the trial would be carried out. In the case of a life-threatening or incapacitating disease, such a trial could not be carried out as it would be unethical.

Another situation leading to a lack of direct comparison could occur in this ideal situation. In many clinical fields, development activity is intense and two (or more) new therapies could start trials at the same time. In this case, the new therapies T3 and T4 would be compared with T2, the reference treatment when the trials began. At the end of the trials, T3 and T4 would be shown to be superior to T2. This would then raise the question of the relative efficacy and safety of T3 and T4. In absolute terms, there is nothing to prevent a trial of T3 versus T4 being carried out (as this would not imply a possible “loss of chance” for patients). In practice, such inter-sponsor trials are rarely carried out.

The situation becomes even more complex with the use of non-inferiority trials. New therapies can avoid positioning themselves as superior to the reference treatment in order to avoid the situation where T3 fails to demonstrate its superiority over T2 and so finds itself virtually ignored. A trial designed to show that T3 is not inferior ("equivalent") to T2 is carried out rather than a trial of the superiority of T2 versus T3, in the hope that the conclusion will show that T3 is as effective as T2 and that they are therefore interchangeable. We know that this conclusion is false in a way as non-inferiority trial methodology introduces an accepted loss of efficacy to make it possible to reach a conclusion of non-inferiority. At the end of a conclusive non-inferiority trial, T3 is potentially inferior to T2 and in practice, T3 frequently obtains an indication as second-line therapy[4]. The next question is to work out the position of T3 in relation to T1, knowing that this might involve direct comparison with a treatment known to be not the most effective (T1). In theory, one should be able to confirm that T3 is at least as effective as T1 because the non-inferiority margin would have been chosen in order to ensure this (see ICH E10), but an empirical study of the field of non-inferiority trials shows that this situation is very rare (5,6).

2.2 Predilection for placebo-controlled trials

In many fields, use of placebo-controlled trials only remains acceptable during the development of a new therapy even after a first effective treatment has become available on the market (7). For example, this is frequently the case for symptomatic therapies for diseases that are not life-threatening or incapacitating. In these areas, the ideal procedure described above is not followed.

At the end of this type of development, an evaluation may find n comparisons against placebo without any idea of the efficacy of treatments relative to each other (if the evaluation is restricted to only the trials carried out). This situation is very common.

This lack of trials comparing active treatments with each other has been regretted many times both in France and in United States.

An editorial in the New York Times on 16 November 2003 under the headline "Head-to-Head Drug Combat"[5] reported that "... For the most part, drugs in this country are not tested against other drugs in the same class. Instead they are tested against a placebo, and if shown to be comparatively safe and effective are approved for marketing. That leaves both patients and their doctors uncertain which approved drugs are better than their competitors and whether high-priced drugs warrant their added cost compared with lower-cost alternatives".

In France, this deficiency has also been mentioned in a report from the *Cour des Comptes* (Court of Auditors) on the application of Social Security funding legislation published on 12 September 2007, Chapter 9 (8):
"In addition, it is a problem that clinical trials against comparators are not compulsory and are therefore not routinely carried out. The Transparency Commission does not have an exact figure for the proportion of dossiers that include clinical trials against comparators, but an estimate based on a sample suggests that fewer than half of all dossiers contain these data."

2.3 Non-transitivity of statistical tests

Another problem likely to arise in any attempt to set up a hierarchy in the efficacy (or safety) of multiple therapies is the fact that certain statistical tests used in clinical trials are non-transitive (9-11).

This means that even if all the direct comparisons needed are carried out, the conclusions of the different trials could contradict each other, making it impossible to classify the therapies.

[4] In this context, second-line therapy means e.g. therapy to be used if a first-line therapy is contraindicated, or if the patient is unable to tolerate it.

[5] http://query.nytimes.com/gst/fullpage.html?res=9A0CE1DE1338F935A25752C1A9659C8B63, viewed on 2 Nov. 2007

For example, in a single three-arm trial comparing therapies A, B and C, it is possible to find that A is better than B, B is better than C and C is better than A, because certain tests are non-transitive[6].
Non-transitivity is well known for Wilcoxon's test, but it also occurs in the whole range of tests based on rank, e.g. the logrank test and the Cox model, which are used very widely in clinical trials.
So this phenomenon restricts any approach based solely on head-to-head trials, leading to 2 by 2 comparisons of therapies. The paradoxes resulting from non-transitivity of tests are resolved by using methods involving global estimation of the network of trials, described in Section 4.4.

2.4 Non-inferiority clinical trials

Non-inferiority trials are common. They are set out to demonstrate that a new therapy is not inferior to a standard therapy. This conclusion is based on an accepted potential loss of efficacy. It means that a conclusive non-inferiority trial does not provide any conclusion on whether the new therapy is as effective as the reference therapy. It can only be concluded, with a confidence level of 5%[7], that it does not lead to a loss of efficacy greater than the margin chosen. A confidence level limited to 5% can allow to this conclusion even when the new therapy is markedly inferior to the reference therapy. In view of this accepted loss of efficacy, the new therapy has to offer advantages over the standard treatment in areas other than efficacy (safety, administration route, etc.).

In practice, non-inferiority margins are still chosen arbitrarily and do not comply with the regulatory requirements of ICH E9 (Section 3.3.2). The margin chosen should ensure that the loss of efficacy caused by the new therapy will not be greater than the efficacy of the standard therapy, otherwise the non-inferiority trial could be conclusive with a new therapy that is inferior to the standard therapy: "An equivalence margin should be specified in the protocol; this margin is the largest difference that can be judged as being clinically acceptable and should be smaller than differences observed in superiority trials of the active comparator".

As interpretation of the results of non-inferiority studies should not be intuitive and verification of the acceptability of the margin chosen requires that results from trials carried out with the standard treatment be available, it has been proposed a number of times that non-inferiority trial results be transformed into a comparison against a putative placebo using an extrapolation process similar to that used in indirect comparisons.

An approach based on a network of trials can incorporate both non-inferiority and superiority trials and so unify interpretation of the results of these different types of trials, without taking into account the non-inferiority margins used (which very frequently cannot be justified).

---

[6] Readers interested in this statistical curiosity can read more about it on the internet: http://en.wikipedia.org/wiki/Nontransitive_dice, http://mathworld.wolfram.com/EfronsDice.html.

[7] In fact 2.5%, but similar to the confidence level of a superiority trial to conclude in favour of superiority, due to the sidedness of the hypotheses leading to the recommendation of the new treatment (see www.spc.univ-lyon1.fr/polycop Sections: *test unilatéraux et essais de non infériorité* [one-sided tests and non-inferiority trials])

3. Literature review methods and selection criteria

3.1 Information sources

3.1.1 Publication databases searched

- Biosis Previews (BioScience Information Service, United States);
- Current Contents Search (Institute for Scientific Information - Thomson Scientific, United States);
- Embase (Elsevier, Netherlands);
- Medline (National Library of Medicine, United States);
- Pascal (*Institut national de l'information scientifique et technique*, France).

3.1.2 Other sources

- Cochrane Library (Wiley Interscience, United States);
- Cochrane Methodology Register*;*
- Website of organisations publishing guidelines and health technology assessments (see Annex 3);
- Google Scholar;
- Lists of references for documents selected.

3.2 Literature search strategy

Publications databases were searched for the following periods:

- Embase, Medline and Pascal: January 1985 to April 2009;
- Biosis Previews: January 1993 to September 2007;
- Current Contents Search: January 1990 to September 2007;
- Cochrane Methodology Register: Issue 2, 2007.

The publication databases searched currently contain no thesaurus terms for the concept of "indirect comparison", so only free text searches were performed. Free text terms were combined in as many stages as required, using the operators "AND" "OR" "EXCEPT".

Figure 2 illustrates the publication database search strategy. Only publications in French or English were searched for.

Figure 2 - Search strategy in the Biosis Previews, Current Content Search, Embase, Medline and Pascal databases

indirect comparison*/title OR *comparaison*indirecte**/title OR indirect analysis/title OR *analyse* indirecte**/title OR
[(indirect evaluation*/title AND direct/title) OR (*evaluation** *indirecte**/title AND *direct**/title)]

OR

[indirect comparison*/title, abstract OR *comparaison indirecte*/title, abstract OR indirect analysis/title, abstract OR *analyse indirecte*/title, abstract OR indirect evaluation*/title OR *evaluation indirecte*/title, abstract OR (indirect evidence*/title AND direct/title) OR indirect treatment* comparison*/title, abstract OR mixed treatment*/title OR multiple treatment*/title OR several treatment*/title OR mixed treatment* comparison*/title, abstract OR multiple treatment* comparison*/title, abstract OR several treatment* comparison*/title, abstract] AND [randomized controlled trial*/title, abstract OR *essai randomisé*/title, abstract OR randomized controlled trial*/term OR *essai randomisé*/term OR meta-analys*/title, abstract OR metaanalys*/title, abstract OR meta-analysis/term OR *metaanalyse*/term]

* truncation

3.3 Publication selection criteria

Publications were selected which proposed indirect comparison methods, carried out studies of the validity of indirect comparisons or discussed the benefits and limitations of these approaches, together with editorials and procedural or position statements issued by national and international health agencies. A few examples of the use of indirect comparisons were also selected.

4. Various methods used to carry out indirect comparisons

In general, any method that produces an estimate of the effect of A compared with B (relative risk, odds ratio, difference in risk, hazard ratio, difference in mean) based on estimates of the effect of A versus control and of B versus control, can be used to carry out an indirect comparison.

Three main types of method are described in this chapter:

- invalid and inappropriate methods;
- the adjusted indirect comparison method, comparing treatments 2 by 2;
- methods simultaneously modelling all competing treatments used in the pathology studied.

4.1 Inappropriate methods

There are two methods, regularly reported in publications, that are not suitable for carrying out valid indirect comparisons. No specific terminology currently exists for describing these methods. We propose to identify them by describing their principle, i.e. “naive comparisons” of point estimates or of active arms from separate controlled trials.

4.1.1 Naive comparison of point estimates

This approach, which consists of comparing point estimates of the efficacy of A versus placebo[8] and that of B versus placebo, is inappropriate. It may appear as natural way of thinking while comparing two active treatments but unfortunately such approach has many disadvantages.

As an example, we may compare the relative risk of mortality obtained during evaluation of A versus placebo to that obtained during evaluation of B versus placebo. In theory, any other index of efficacy (relative risk reduction, odds ratio, relative reduction in odds, difference in risk, number of patients to be treated (*Number Needed to Treat “NNT”*), absolute or relative difference in means, etc) can be used instead of relative risk. The treatment giving the greatest reduction in relative risk will be classified as more effective than the other.
For example, two treatments, A and B, are being studied in mortality trials, the results of which are shown in Table 3.

Table 3 - Results for point estimates from trial A versus placebo and trial B versus placebo

	Relative risk All-cause death
Treatment A (versus placebo)	0.76
Treatment B (versus placebo)	0.87

Treatment A resulted in a 24% relative reduction in mortality, which is higher than treatment B (13%). Treatment A will therefore be classified as "more effective".
Apart from any considerations on the methodology used in the trials, populations included and treatments delivered, the main disadvantage of this approach is that the statistical precision of the estimates is not taken into account. In fact it is quite likely that the most favourable estimate will also be the one with the largest confidence interval, so that the least favourable limit of the confidence interval is compatible with low efficacy. In this case, a conclusion based on point estimates will favour the treatment whose guaranteed minimum benefit[9] is lowest, to the detriment of the treatment which assures us that it provides superior benefit, even in the worst case of statistical error during estimation. This situation is illustrated in Table 4.

[8] To make it easier to describe, we will refer to comparisons versus placebo. However, all the comments apply irrespective of the comparator used, i.e. placebo, standard care, active therapy.

[9] The guaranteed minimum benefit (at 95%) is the interpretation of the least favourable limit of the confidence interval.

Table 4 - Results of point estimates of trial A versus placebo and of trial B versus placebo

	Relative risk All-cause death (95% CI)
Treatment A (versus placebo)	0.76 (0.53-0.99)
Treatment B (versus placebo)	0.86 (0.84-0.90)

With treatment B, a reduction in mortality of at least 10% is guaranteed with a 95% degree of certainty, while with A we have the assurance of a 1% reduction (with a 95% degree of certainty). An actual example of the use of this approach is given in a case study in Section 9.3.1.

This examination of the least favourable limit of the confidence interval illustrates that comparison of point estimates is inappropriate. It does not yield a satisfactory indirect comparison method: it focuses on a single very specific option (the least favourable limit of the confidence interval) and ignores the range of other options offered by the confidence interval. Other methods that have been developed circumvent this problem.

In addition, carrying out this type of comparison with differences in risk or in NNT is distorted by possible inter-trial differences in baseline risk.

4.1.2 Naive comparison of active arms

Another inappropriate approach consists of comparing the active arms of trials without taking account of the randomisation, i.e. pairing within the trial.
Based on two trials of A versus placebo and B versus placebo, the only data used are those from arm A of the first trial and from arm B of the second trial. The two placebo arms are effectively put aside. The effect of A versus B is then estimated by comparing the results of arm A with those of arm B. As a result, confounding factors specific to each study are not controlled for. There is a total loss of the controlled nature of both trials. The result obtained is no longer free of confounding bias. Similarly, the benefit of randomisation is totally lost and the comparison obtained is not free of selection bias. In fact, although this approach is based on two randomised controlled trials, it no longer has any methodological reliability and becomes a purely observational study.

Based on the results given in Table 5 and Table 6, this approach would compare 45/1 567 to 36/1 251, i.e. 2.87% to 2.88%. This comparison would suggest that A cannot be considered more effective than B. This conclusion is paradoxal as examination of the two trials shows that A was shown to be effective against placebo, while B was not shown to be effective. The proposed comparison of A versus B is confounded by the basic risk.

Table 5 - Results of trial comparing A with placebo

	Treatment A	Placebo	Relative risk (95% CI) p-value
Deaths	45/1 567 (2.87%)	103/1 546 (6.66%)	0.43 (0.31-0.61) p<0.001

Table 6 - Results of trial comparing B with placebo

	Treatment B	Placebo	Relative risk (95% CI) p-value
Deaths	36/1 251 (2.88%)	42/1 298 (3.24%)	0.89 (0.57-1.38) p=0.6

In addition to the theoretical arguments against the validity of this approach (see above), its biased nature has been demonstrated empirically (12). The case study in Section 9.3.2 describes a real example where this method has been used.

Although this method is methodologically completely invalid, it is one of the sources of current mistrust with regard to indirect comparisons. It was the first method to be used, but the problem it raises is not present in all indirect comparisons. The methods developed specially provide a solution to this problem, as will be shown in Section 4.2.1.

4.2 Adjusted indirect comparisons

Adjusted indirect comparison is the first valid method available that makes it possible to perform indirect comparisons.

4.2.1 Principle

This two by two indirect comparison method incorporates an estimate of the effect of A versus placebo and of the effect of B versus placebo. So it preserves the randomisation and retains the methodological properties of the randomised controlled trial.

The calculation required is very simple. As an example, we could consider a case in which the endpoint of interest is binary and in which the effect is measured by relative risk.
Let RR_a be the relative risk obtained by comparing treatment A with placebo and RR_b be the relative risk obtained by comparing treatment B with placebo. The result of indirect comparison of A versus B is the relative risk $RR_{a/b}$, obtained from

$$RR_{a/b} = \frac{RR_a}{RR_b} \tag{0.1}$$

The confidence interval of this relative risk is calculated from the confidence interval of its logarithm, using as variance the sum of the variances of the logarithms of the two relative risks.

$$\mathrm{var}\left(\log RR_{a/b}\right) = \mathrm{var}\left(\log RR_a\right) + \mathrm{var}\left(\log RR_b\right)$$

Based on this variance, the confidence interval of the relative risk $RR_{a/b}$ is obtained in a standard way.

A similar calculation makes it possible to extrapolate the benefit compared with placebo of a new treatment, based on the RR of this treatment compared with the reference treatment and the RR of the reference treatment compared with placebo. This calculation is useful for non-inferiority trials.

It is very easy to demonstrate the expression (0.1). In a direct comparison trial, the risk of an event in patients without treatment is r_0. In the arm receiving A, the risk under treatment r_a becomes

$$r_a = r_0 \times RR_a \tag{0.2}$$

where as before, RR_a stands for the effects of A compared with "no treatment" (measured in a comparison of A versus placebo).
In the same way, the risk under treatment r_b in the treatment B arm is

$$r_b = r_0 \times RR_b \tag{0.3}$$

The effect of A compared with B in this trial is measured by the relative risk $RR_{a/b}$ which by definition is

$$RR_{a/b} = \frac{r_a}{r_b}$$

which becomes by using expressions (0.2) and (0.3)

$$RR_{a/b} = \frac{r_a}{r_b} = \frac{r_0 RR_a}{r_0 RR_b}$$

$$= \frac{RR_a}{RR_b}$$

which justifies calculation of the adjusted indirect comparison.
In these calculations, the basic risk r_0 disappears completely. This extrapolation based on measures of effects is therefore completely insensitive to the basic risk of the trials of A and B. The indirect comparison is therefore valid even though the patients studied in the trials of A and B are not the same. The condition of validity is found elsewhere, as described in the next section.

4.2.2 Condition of validity: stability of effects

For this method to produce a clinically meaningful result, there has to be some consistency between the two relative risks RR_a and RR_b. In other words, under the same conditions (criteria of severity of disease, patient characteristics, concomitant therapy, etc) as those used to evaluate treatment B against placebo, evaluation of treatment A would produce the same relative risk as that observed in the trial of A versus placebo which was actually carried out (the same should also be verified for treatment B).
In other words, this condition implies that placebo-controlled trials of A and B are similar in terms of interaction of covariables[10] with treatments, in other words that there is a stability of effects for the different situations in which the trials included in the calculations were carried out (13).

An interaction covariable is a variable related to patients, the context of the trial or the methods of administration which affect the efficacy of treatment, the latter being expressed by variation in the relative risk according to the value of this covariable. These interactions are identified from subgroup analysis, and more particularly during an interaction test.
Table 7 gives an example of an interaction. The effect of treatment A is probably not the same in diabetics and in non-diabetics. So the global relative risk of the trial depends on the distribution of patients between diabetics and non-diabetics.
If the proportion of diabetics in a trial in which treatment B is evaluated is greater than the proportion of diabetics in a trial of treatment A, a comparison of A versus placebo under the conditions of the trial of B will not give the same relative risk as that obtained based on the actual trial of A.

Table 7 - Example of interaction between the effect of treatment A (measured here by relative risk) and a patient characteristic (diabetic or non-diabetic)

	Population	Relative risk Treatment A versus placebo	Interaction
Global trial	1 230	0.90	
Diabetics	430	0.76	p=0.001
Non-diabetics	800	0.95	

Establishing the relationship between the two relative risks then raises the problem of knowing what type of patients the $RR_{a/b}$ obtained represents; it is certainly not representative of patients included in trial B, as A would not have provided the relative risk used in the calculations involving these patients (there were fewer diabetics in the trial providing the estimated RR_a).
If treatment B is not subject to any interaction (i.e. there are no variables modifying its effect), its comparison against placebo under the conditions of the trial of treatment A would have given the same risk as that obtained. So the ratio RR_a/RR_b is valid for the population studied in the trial of A but not for the population of patients included in the trial of B.
If these two treatments are subject to interactions, the relationship is no longer clinically meaningful.
In practice, eliminating this type of problem requires verifying for both treatments that the effect of treatment is not being substantially modified by any covariables. This is done by examining the interaction tests in the subgroup analyses. However, the diagnosis may be made difficult by a lack of power of the interaction tests and by the possibility that the relevant variables have not been analysed.

In contrast, patients in the placebo-controlled trials of A and B are not necessarily identical except for interaction covariables. In fact, although they differ in their risk of events under placebo, this difference would not modify RR_a or RR_b since the relative risk performs an abstraction of the level of basic risk by calculating a ratio.

A problem arises when variables change the relative risk. So the trials of A and of B have to be similar at the level of interaction covariables. The problem is due to the fact that the interaction covariables have not been clearly identified. It is necessary to rely on the subgroup analyses performed in the trials to identify any interaction covariables that may be present and, if necessary, to verify whether the distribution of these covariables is the same in all trials using the table of baseline characteristics.

[10] Here, the term "interaction" is used in the sense in which it is used in clinical trials and not in the sense of pharmacological interactions when drugs are used in combination. It refers to variables which modify the value of the treatment effect, for example of the relative risk. This interaction is explored in trials using e.g. subgroup analyses.

However, it is quite rare for strong interaction covariables to be present. It is fairly likely that the conditions of validity will have been satisfied if the trials of A and B were carried out under similar conditions (trials carried out at the same time, availability of similar concomitant treatments) and in patients with similar characteristics.

4.2.3 Use of meta-analysis

In practice, there may be a number of placebo-controlled trials of A and of B. Indirect comparison calculations require a single relative risk for A and a single relative risk for B. Selecting a single trial for A and a single trial for B would introduce an unacceptable arbitrary factor into the process. This is resolved by first carrying out a meta-analysis of all placebo-controlled trials of A and a meta-analysis of all placebo-controlled trials of B. As the meta-analysis summarises all the information from several trials into a single result, at the end of this preliminary stage there will be only two relative risks (Figure 3).

In addition, the fact that results from a meta-analysis can be applied more generally makes them superior to results obtained from a single trial, as meta-analyses group together trials carried out in different populations according to different protocols. This summarising of the data helps to increase the stability of the effect of treatments and most importantly, makes it possible to test this stability.

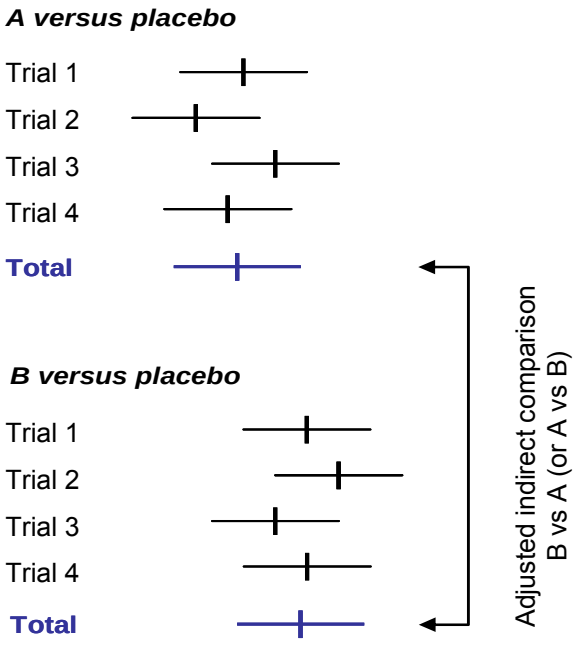


Figure 3 - Use of meta-analysis to calculate adjusted indirect comparisons.

4.2.4 Possibility of multiple paths and evaluation of inconsistency

In some cases, adjusted indirect comparisons could be carried out using a number of paths, rather than just placebo.
The term "path" actually stands for the common comparator of the basic comparisons on which the extrapolation performed by the adjusted indirect comparison is based.
In the simple case we have described so far, the indirect comparison of A versus B is carried out using the path A→ placebo → B as both basic comparisons used were carried out versus placebo.
An indirect comparison of A and B could also be based on a different path if trials of A and of B versus another common comparator were available, e.g. an active treatment, C. The availability of trials of A versus C and of B versus C would make it possible to estimate A versus B by the path A→ C → B.
Calculations would therefore provide two independent estimates of the indirect comparison of A versus B (the one using placebo and the one using C). If both these results are available it will be possible to test whether there is any inconsistency between the two paths, i.e. whether the result of the indirect comparison varies depending on the path (Figure 4). This test is based on the heterogeneity between the two estimates produced of the relative risk A vs. B, i.e. $RR_{A/B}$ and $RR'_{A/B}$.
If there is any heterogeneity (i.e. if there is a statistically significant difference between $RR_{A/B}$ and $RR'_{A/B}$), there is inconsistency: the result varies according to the path. This situation is problematic because the result should be the same irrespective of the path. The estimates should be interpreted very cautiously, unless there is a clear and non-arbitrary explanation for the inconsistency.
If there is no heterogeneity, the final result will be a meta-analysis of the two estimates $RR_{A/B}$ and $RR'_{A/B}$ giving the highest possible level of accuracy. If no inconsistency is detected, greater confidence may be placed in the result.

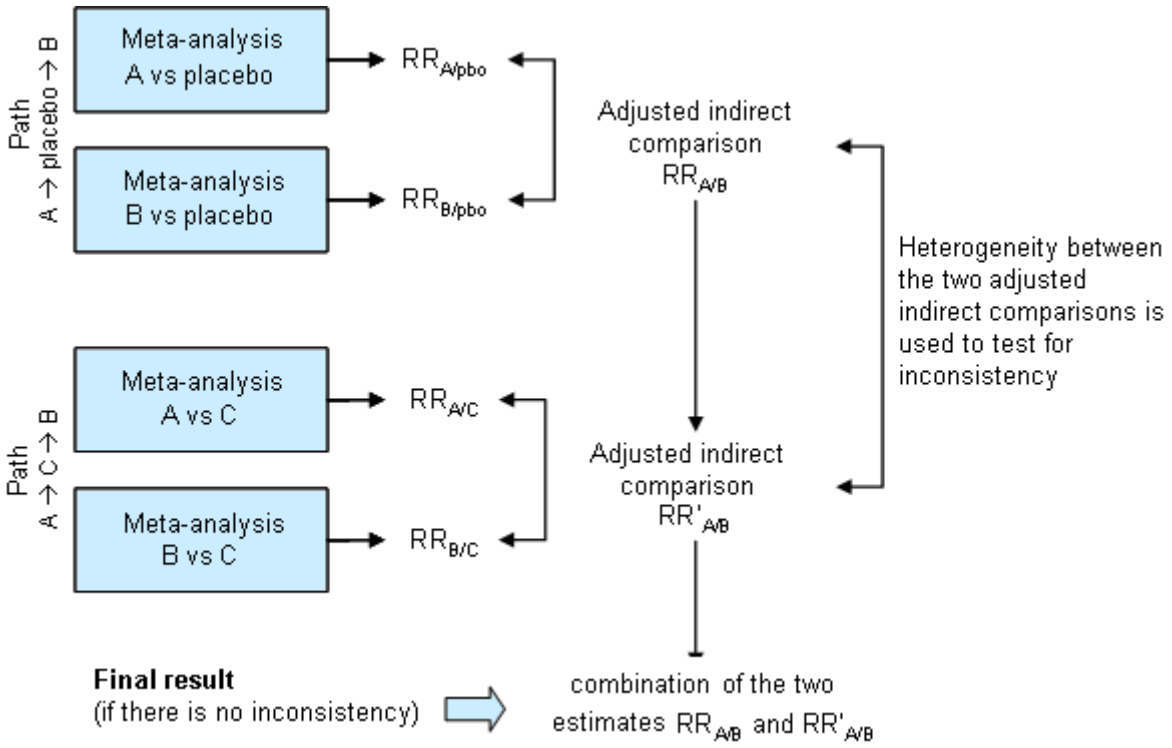


Figure 4 - Illustration of the adjusted indirect comparison process if there are a number of basic comparison paths

4.2.5 Comparisons of more than two treatments

Evaluations are often performed to compare more than two competing treatments simultaneously.
Using the adjusted indirect comparison method, comparison of n treatments will mean proceeding in two by two steps, as adjusted indirect comparisons only consider two treatments at a time.
For four treatments T1, T2, T3 and T4, this means making six comparisons:

- T1 vs. T2;
- T1 vs. T3;
- T1 vs. T4;
- T2 vs. T3;
- T2 vs. T4;
- T3 vs. T4.

The problem of multiple comparisons arises, leading to significance level inflation. Under the null hypothesis that all treatments have the same efficacy (i.e. that there is no difference between treatments when considered two by two), there are still six ways of finding a difference due to chance alone (the six two by two comparisons). If each unit comparison is carried out with a significance level of 5%, the global risk of incorrectly finding at least one difference among the four treatments is substantially increased, rising to 26%. We have a one in four chance of concluding that one treatment is more effective than the others when in reality they are all equally effective.
So adjusted indirect comparisons have the advantage of being easy to use and easy to understand. However, the limitations of the method soon appear when the question to be answered covers more than two treatments. Other methods have been developed that do not have these limitations in this situation. These are methods based on modelling, such as network meta-analysis and meta-regression.

4.3 Mixed approach for direct and indirect comparisons

4.3.1 Principles

Indirect comparisons can still be useful even if direct comparison trials are available.
Often only one direct comparison trial is available. Quite often this trial has been designed with a lack of power. In other cases, the comparator may have been used in ways which are debatable. In such a situation a mixed approach, called a mixed treatment comparison[11], in which the results of direct comparisons are compared with those of indirect comparisons, is very useful as it removes or confirms any reservations that one might have about direct comparison trials (Figure 5).

The mixed approach can also be used to assess the validity of the results of an adjusted indirect comparison.
If the estimates derived from direct and indirect comparisons are in accordance with each other, the results are combined together to give a global estimate of the difference of effect between A and B, taking account of all the information available.

[11] At present there is no French term for this concept.

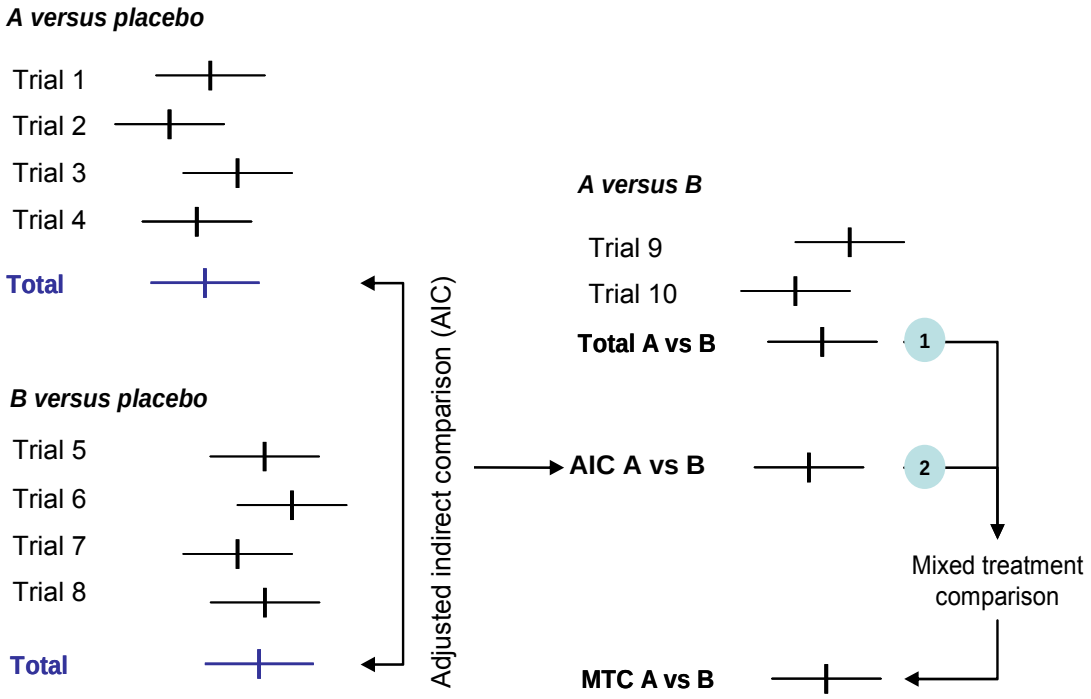
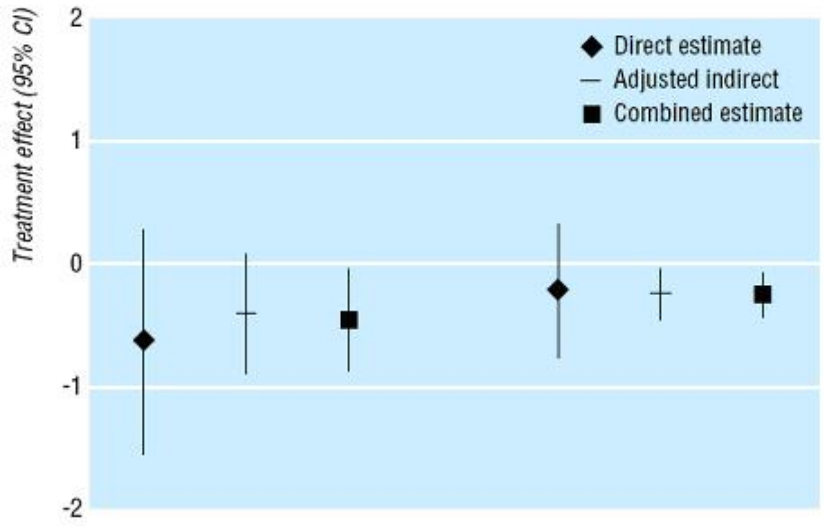


Figure 5 - Illustration of the mixed treatment comparison (MTC) process. The result of the meta-analysis of direct comparison trials: (1) is combined with the result of the meta-analysis of indirect comparisons (2) and the concordance of these two estimates is tested by a test of heterogeneity.

4.3.2 Example

Estimates derived from adjusted indirect comparisons and estimates obtained from direct comparisons can easily be grouped together by meta-analysis.
Song *et al.* (14) give two examples of the application of this method taken from publications by Soo *et al.* (15) and Trindade and Menon (16) where, as the estimate from direct comparisons is completely in line with the estimate from indirect comparisons, the best possible estimate is one that combines the results of the two approaches (Figure 6).



Soo *et al*. (15) Trindade and Menon (16)

Figure 6 - Two examples of integration by a meta-analysis of the results of direct comparisons and the results derived from adjusted indirect comparisons
Reproduced from Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 2003;326(7387):472. Copyright 2003, with permission from BMJ Publishing Group Ltd.

4.4 Methods based on a global estimate from a network of trials

4.4.1 Principle of a network of trials

The general idea of this approach is to represent all the information coming from the different trials in a specific clinical field in the form of a network ("Network of evidence") (17). In this network the points represent treatments and the arrows linking the points represent the available comparisons (provided by trials comparing treatments situated at either end of the arrow, the direction of the arrow giving the direction of the comparison and the relative risks). Figure 7 illustrates an example of such a network derived from a meta-analysis of antihypertensive therapies published in 2007 by Elliott and Meyer (18). This meta-analysis investigated onset of diabetes after antihypertensive therapy.

The global estimation methods proposed (see Section 4.4.4) make it possible to assess the efficacy of all treatments compared with a reference treatment which may or may not have been chosen arbitrarily. The results are shown in the form of classical meta-analysis graphs which are instantly comprehensible.

The estimate is global, using existing both direct comparisons and the indirect comparisons that can be made. A hierarchy is established quantitatively from the estimate of size of effect, which protects it from the problems related to non-transitivity of statistical tests (see Section 2.3).

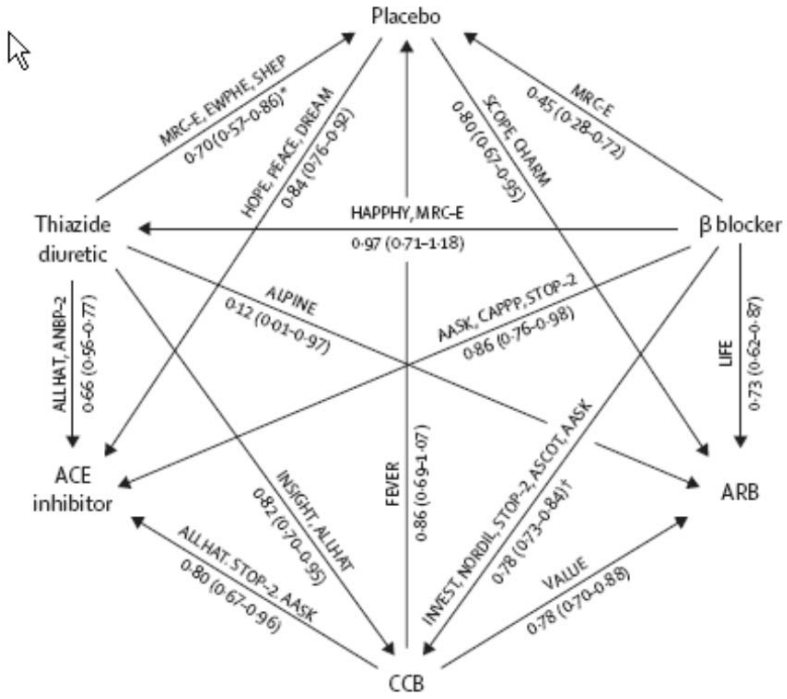


Figure 7 - Network of trials of antihypertensives which reported incidence of diabetes. The direct comparisons available are represented by arrows labelled with the name of the trial, odds ratio and confidence interval according to Elliott and Meyer, 2007 (18)

Reprinted from Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. Lancet 2007;369(9557):201-7. Copyright 2007, with permission from Elsevier

NB: interpreting the results to find the treatment associated with the lowest incidence of diabetes means making a naive indirect comparison of odds ratios. At the end of this analysis, angiotensin receptor blockers (ARB) were considered to be superior to ACE inhibitors as the odds ratio for angiotensin receptor blockers versus diuretics was better than the odds ratio for ACE inhibitors versus diuretics. However, a correct interpretation of these results should be restricted to "This network meta-analysis shows that placebo, calcium channel blockers (CCB), ACE inhibitors and angiotensin receptor blockers (ARB) cause fewer incident cases of diabetes than diuretics". It is not possible to draw any conclusion about beta-blockers.

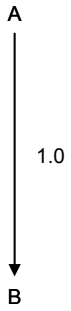
4.4.2 Concept of inconsistency

The simplest network is one with only two treatments, A and B (Figure 8). The value shown next to the arrow is an estimate of the effect of B compared with A. Next we will consider that the effect is estimated from the difference (in risk for example) designated by δ_{AB}:

$$\delta_{AB} = x_B - x_A$$

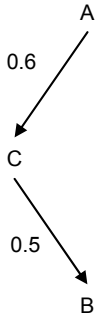
where x_A and x_B stand for values obtained in the A and B arms of a comparison of A versus B. x may be a risk (in this case δ_{AB} is the difference in risks) or a mean (and δ_{AB} is the difference in means).

Figure 8 - Network involving two treatments, A and B



A situation where two comparisons are available, i.e. C compared with A and B compared with C, results in the network shown in Figure 9.

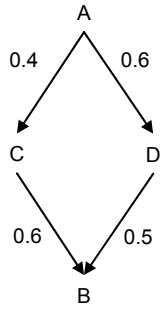
Figure 9 - Network consisting of two types of comparison C versus A and B versus C



A simple way of obtaining the effect of B compared with A is to add $\delta_{AB} = \delta_{AC} + \delta_{CB}$
i.e., with the numerical values obtained from the example in Figure 9, δ_{AB} = 0.6+0.5 = 1.1.

If four types of comparison are available, the network of these results can be illustrated by the diagram shown in Figure 10.

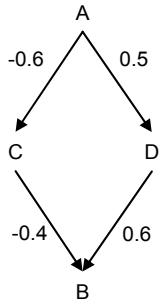
Figure 10 - Network involving 4 treatments A, B, C, D



The same difference δ_{AB} can now be obtained by two different paths: A → C → B and A → D → B. δ_{AB} is obtained by $\delta_{AB} = \delta_{AC} + \delta_{CB}$ for the first path and by $\delta_{AB} = \delta_{AD} + \delta_{DB}$ for the second, i.e. with the numerical values: 0.4 + 0.6 = 1.0 and 0.6 + 0.5 = 1.1. So this results in two estimates of δ_{AB} that are very close to each other irrespective of the comparison path used. There is a certain consistency in the network of possible indirect comparisons.

In the situation illustrated by Figure 11, the first path gives the estimate of δ_{AB} as -0.6-0.4=-1.0 and the second 0.5+0.6=1.1. Here, there is a complete discrepancy between the two estimates, indicating inconsistency in the comparison network. No indirect estimate of A versus B will be possible. So the notion of inconsistency is an important factor in the validity of indirect comparisons (obtained by means of a network of comparisons). If it can be tested, i.e. if a number of paths exist for performing the same extrapolation, the absence of inconsistency is a factor supporting the validity of the extrapolations obtained: the same estimate is obtained irrespective of the procedure used.

Figure 11 - Example of inconsistency in a network of comparisons

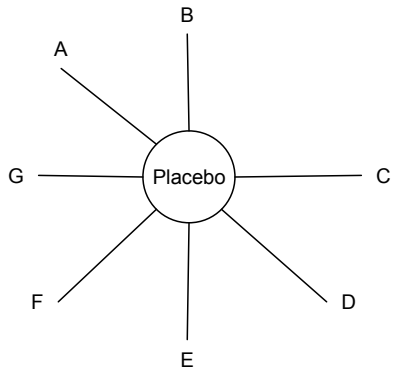


4.4.3 Networks of comparisons corresponding to the different development plans

The different development plans described in Sections 2.1 and 2.2 can be illustrated using a network of comparisons.

The network shown in Figure 12 corresponds to a situation in which all competing treatments have been developed using placebo-controlled trials.

Figure 12 - Network for evaluating all competing treatments against placebo



The network in Figure 13 represents a response to a situation in which a placebo cannot be used once an effective therapy has become available (e.g. evaluation of treatments intended to reduce mortality). As a control, trials use the last treatment which was shown to be superior to a previous one.

Figure 13 - Network in which a placebo cannot be used once an effective treatment has become available

4.4.4 Proposed global estimation methods

Several models have been proposed in the literature to represent networks. Several statistical methods for estimating the parameters for these models (particularly those characterising treatment comparisons) have also been proposed. This diversity has resulted in the following classification:

- Estimation using Bayesian methods: Bayesian network meta analysis
  - Lu and Ades model (19)
  - Model of Caldwell *et al.* (20)
- Estimation using a mixed linear model
  - Lumley's network meta-analysis (21)

4.5 Bayesian network meta-analysis

The most complete of the different models proposed for carrying out Bayesian network meta-analyses appears to be the one proposed by Lu and Ades (19). It follows on from previous models proposed by Caldwell *et al.* (20) (see below) and Smith *et al.* (22). All these models are based on the proposition of Bayesian meta-analysis made by Higgins and Whitehead (23).

These methods are implemented using Markov chain Monte Carlo integration methods using WinBUGS software.

4.5.1 Lu and Ades model

- *Statistical method*

Lu and Ades (19) propose an extension to I treatments of the model of Smith *et al.* (22) for combining direct and indirect comparisons in order to perform mixed treatment comparisons. Their model also naturally takes into account multi-arm studies which raise the issue of the non-independence (correlation) of comparisons performed using a common control. The estimates are obtained by means of a Bayesian approach implemented using Markov chain Monte Carlo methods.
The basic two-treatment model applies only to binary endpoints and is constructed as follows.
The exponents T and C stand for the test treatment T and the control treatment C respectively. The index *k* stands for the k^{th} trial. r is the number of events observed in a group of size n, p is the probability that the event will occur. The effect of treatment δ_k is allowed to vary from one trial to another. This is therefore a random meta-analysis model. The global effect of treatment (in time as meta-parameter) is estimated by d. At the level of each trial, μ_k stands for the mean frequency of events in the T and C groups which is treated as a nuisance parameter, as it is not a parameter of interest.

With these conventions, the model is

$$r_k^T \approx bin\left(p_k^T, n_k^T\right)$$
$$r_k^C \approx \text{bin}\left(p_k^C, n_k^C\right)$$
$$\text{logit}\left(p_k^T\right) = \mu_k + \delta_k / 2$$
$$\text{logit}\left(p_k^C\right) = \mu_k - \delta_k / 2$$
$$\delta_k \approx \text{normal}\left(d, \sigma^2\right)$$

The value of this model is that it is easy to extend to multiarm trials. It is extended to *I* treatments *i* by:

$$r_{ki} \approx \text{bin}\left(p_{ki}, n_{ki}\right)$$
$$\text{logit}(p_{k1}) = \mu_i - \delta_{k2} / I - \delta_{k3} / I - \cdots - \delta_{kI} / I$$
$$\text{logit}(p_{k2}) = \mu_i + (I-1)\delta_{k2} / I - \delta_{k3} / I - \cdots - \delta_{kI} / I$$
$$\text{logit}(p_{kI}) = \mu_i - \delta_{k2} / I - \delta_{k3} / I - \cdots + (I-1)\delta_{kI} / I$$

Box - Markov chain Monte Carlo methods (MCMC)

Markov chain Monte Carlo methods (MCMC) make it possible to obtain the distribution of a parameter of interest based on observed data. To put it simply, the distribution of a random variable gives the probability that a parameter will take a particular value. By observing the distribution it is possible to determine the most probable value, i.e. at the top of the distribution. So the distribution can be assimilated to a histogram.

If the value of the parameter to be estimated is known exactly, the distribution will be a vertical line. A single value has a 100% probability of being the true value of the estimated parameter, all the others have a 0% probability. In practice this is never the case, and the spread of the distribution expresses the lack of precision with which the true value is known.

The distribution of the parameter to be estimated is established from observed data (as with any estimation procedure). However, the cost of using Bayesian methods is that a hypothetical prior probability is needed for the parameter distribution. So the result produced, i.e. the estimated distribution parameter, depends simultaneously on the data and on the chosen prior. These methods give a posterior distribution (i.e. after the observation of data) based on a prior distribution (as it is imagined prior to the trial, before any information has been obtained from the data). The need for a prior distribution for the element that is not known but which one wishes to know through observation is the main problem with this approach. This is an important point, as the result (the estimated posterior distribution) can change considerably depending on the prior chosen.
Fortunately, it is possible to find uninformative priors, i.e. priors which introduce very little information. So these priors do not affect the result. Such uninformative prior distributions appear to say that before analysis (i.e. before the first data have been observed), any value is possible for the true value of a parameter and that it is therefore not justifiable to favour one more than another.

Obtaining a posterior distribution based on the prior distribution uses Bayes' theorem, which is why the method is known as a Bayesian method. Bayes' theorem states that the posterior distribution $f(\theta|d)$ of the parameter θ taking account of the observed data d is

$$f(\theta|d) = \frac{l(d|\theta)\, f_0(\theta)}{g(d)}$$

where $f_0(\theta)$ is the prior distribution of the parameter and $l(d|\theta)$ is the likelihood of the data according to the value of the parameter. The likelihood is the probability of obtaining the data which have actually been observed for a given value of the parameter. $g(d)$ is the unconditional distribution of data (also called marginal distribution of data or distribution of priors).

With MCMC methods, transformation using information contributed by data from a prior distribution to obtain a posterior distribution will not be done by an analytical process (using a formula) but will be based on intensive calculation using a computer. The value of the parameter θ of the model is estimated by repeating the same process in a loop a very large number of times until a stable estimate of the distribution is obtained.
The estimation process uses random number generation to produce a posterior distribution from a prior distribution of data. The posterior distribution obtained becomes the prior distribution for next loop, and so on. A theorem ensures that the process converges on the true posterior distribution of the parameter to be estimated.

So the process produces a sequence (chain) of numbers which, after the first values have been excluded (burn-in), represents a random sample of the posterior distribution. The distribution of the parameter of interest can be estimated empirically by a descriptive analysis of the sequence (mean, median, quantile, representation of the histogram).

In conclusion, the input for these methods consists of the data observed, a prior distribution and the first value of the parameter. The output is a long sequence of numbers (Markov chain) whose final section is an empirical estimation of the distribution. Assurance that the sequence obtained is a correct random sample of the posterior distribution of the parameter depends on the initial value and the prior distribution chosen, together with the convergence (relative stability) of the sequence obtained. An undiagnosed absence of convergence and/or a poor choice of prior or of initial value leads to incorrect estimates.

4.5.2 Model of Caldwell *et al.*

- *Statistical method*

The model proposed by Caldwell *et al.* (20) is a generalisation of the model proposed by Higgins and Whitehead (23).

This is a meta-analysis in which *k* treatments have been compared with each other in trials. For each trial *j*, let r_{jk} be the number of events observed in treatment arm *k* whose population is n_{jk}.
The number of developments follows a binomial distribution

$$r_{jk} \approx B\left(p_{jk}, n_{jk}\right)$$

in which p_{jk} is the frequency (risk) of onset of the event, an endpoint.
The use of an odds ratio to represent treatment effects involves modelling of logits as follows:

$\text{logit}\left(p_{jk}\right) = \mu_{jb}$ when *k*=*b*, *b* standing for the reference treatment of the comparisons

and $\text{logit}\left(p_{jk}\right) = \mu_{jb} + \delta_{jbk}$ when *k* is different from *b*.

μ_{jb} represents a baseline odds value for the endpoint in the trial j.

This model can be one of two variants based either on a fixed effects model or on a random effects model. Estimates are obtained by MCMC using e.g. WinBUGS software, and require the introduction of uninformative prior distributions into the estimation process.

As an example, this approach was used to compare thrombolytics in the acute phase of myocardial infarction (20).

4.6 Estimation using a mixed linear model

- *Statistical method*

The use of a mixed linear model for performing indirect comparisons by means of a network of comparisons was proposed in 2002 by Lumley (21).

The estimated treatment effect based on trial *k* comparing treatments *i* and *j* is written Y_{ijk} with estimated variance σ_{ijk}^{2}.

The model of the effect Y_{ijk} consists of three components in addition to sampling error:

1. the meta true effect of treatments *i* and *j* which is represented by μ_i and μ_j respectively. The meta true effect is the value around which the true effects of treatment established in each trial fluctuate. The observed effect is derived from the true effect more or less skewed by sampling error (related to sampling fluctuations).

2. the random effect η_{ik} and η_{jk} which represents the difference between the true effect of treatment for trial *k* and the corresponding meta-values. The variance of the 2 effects η_{ik} and

η_{jk} is τ^2. The true treatment effect *i* in the trial *k* is therefore $\mu_i + \eta_{ik}$. Y_{ijk} is the estimate of the difference between the treatments *i* and *j* in the trial *k*, and is therefore an estimate of the difference between the true effects $\mu_i + \eta_{ik}$ and $\mu_j + \eta_{jk}$.

3. a second random effect ξ_{ij} representing the change in the effect of treatment i when compared with j. ξ_{ij} therefore makes it possible to capture any inconsistency in the network of evidence. Its variance ω makes it possible to measure the inconsistency of the network.

So in conclusion, the model is

$$Y_{ijk} = \left(\mu_i + \eta_{ik}\right) - \left(\mu_j + \eta_{jk}\right) + \xi_{ij} + \varepsilon_{ijk}$$
$$\varepsilon_{ijk} = N\left(0, {\sigma_{ijk}}^2\right)$$
$$\eta_{ik} \approx N\left(0, \tau^2\right)$$
$$\xi_{ij} \approx N\left(0, \omega^2\right)$$

Where ε_{ijk} stands for random sampling error which at the level of trial *k* skews estimation of the difference between treatment *i* and treatment *j*.

Box - Mixed model, hierarchical

In statistics, an effect is the change induced by a factor in the value of a random variable. In a trial, the treatment factor causes a change in the value of the variable used as endpoint. For most of the time, factors are regarded as fixed. The treatment effect is therefore the mean constant change induced by treatment on the random variable used as an endpoint.

Mixed effects models, also called random effects models, are used to take account of several levels of fluctuation/random errors (random effects) and so to consider that the effects may also vary in a random manner.
Qualification of an effect as random means that this effect varies from one statistical unit to another, in a random manner according to a certain distribution.

In a meta-analysis, modelling using a fixed model represents reality by considering that the effect *Y* observed in the trial *k* is equal to the true treatment effect, plus the random fluctuations of sampling (i.e. the random error occurring at the level of the trial between the observed value and the true value).
So in this model there is only one effect (the true effect of treatment) and this effect is fixed for all trials (hence the name "fixed effects model").

A random meta-analysis model will stipulate that the true effect of treatment varies from one trial to another, fluctuating at random around a value which is a meta-parameter. The effect being investigated at the level of the trial is therefore not fixed but random, hence the name "random effects model". The meta-analysis then tries to estimate the meta-parameter. So the model stipulates heterogeneity of effect of treatment (the effect varies from one trial to another at random).
In this random model, we therefore have two levels of variability due to chance:
1. sampling fluctuations (as with the fixed model) describing the difference between the true value of the effect of treatment in a trial and the effect observed
2. fluctuations of the true effect between trials.

- *Example of application*

This method was used in a meta-analysis by Psaty *et al.* of first-line antihypertensives (24). The objective was to complete and estimate all possible comparisons in the network. For example, there were no trials directly comparing low-dose diuretics with angiotensin receptor blockers. An indirect comparison performed implicitly by the method made it possible to document the relative efficacy of these two types of treatment (Figure 14 and Figure 15).

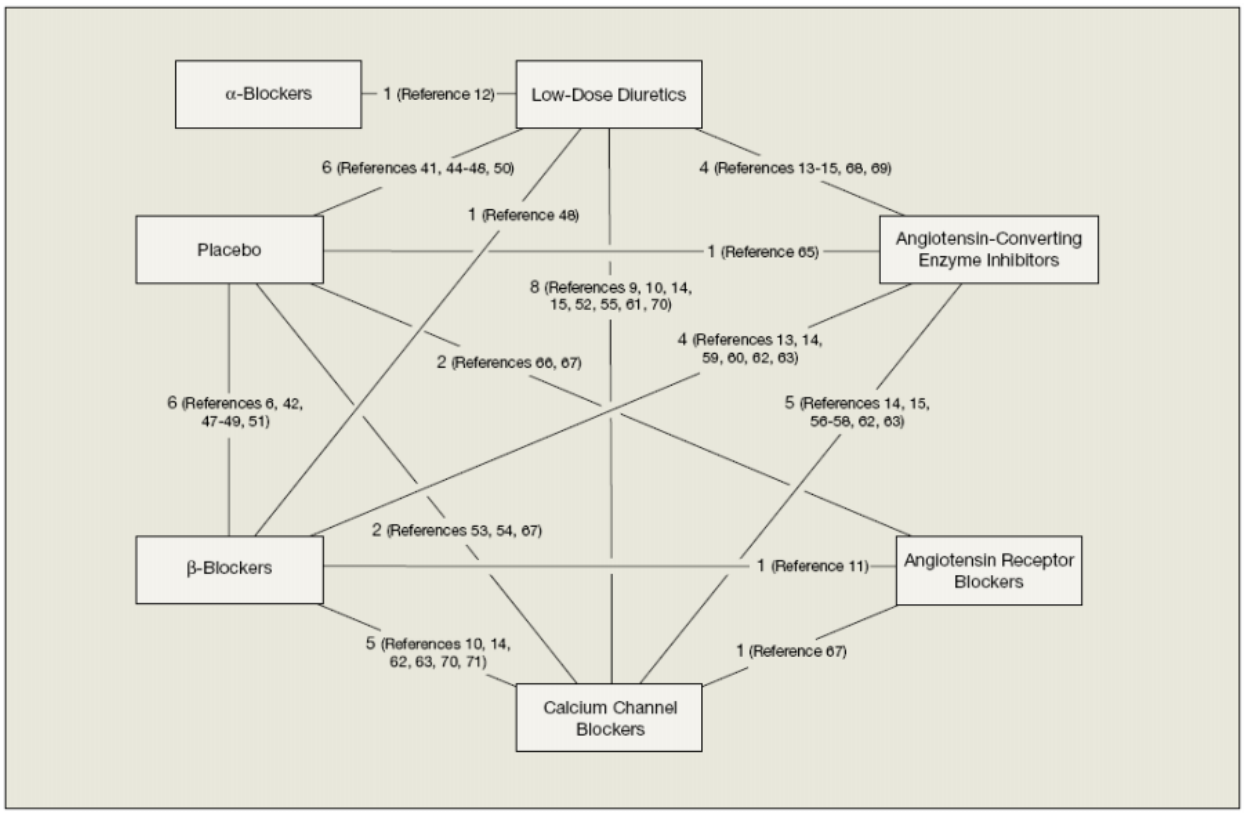


Figure 14 - Network of comparisons included

Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, *et al.* Health outcomes associated with various antihypertensive therapies used as first-line agents. A network meta-analysis. JAMA 2003;289(19):2534-44.

Figure 2. Network Meta-analysis of First-Line Treatment Strategies in Randomized Controlled Clinical Trials in Hypertension

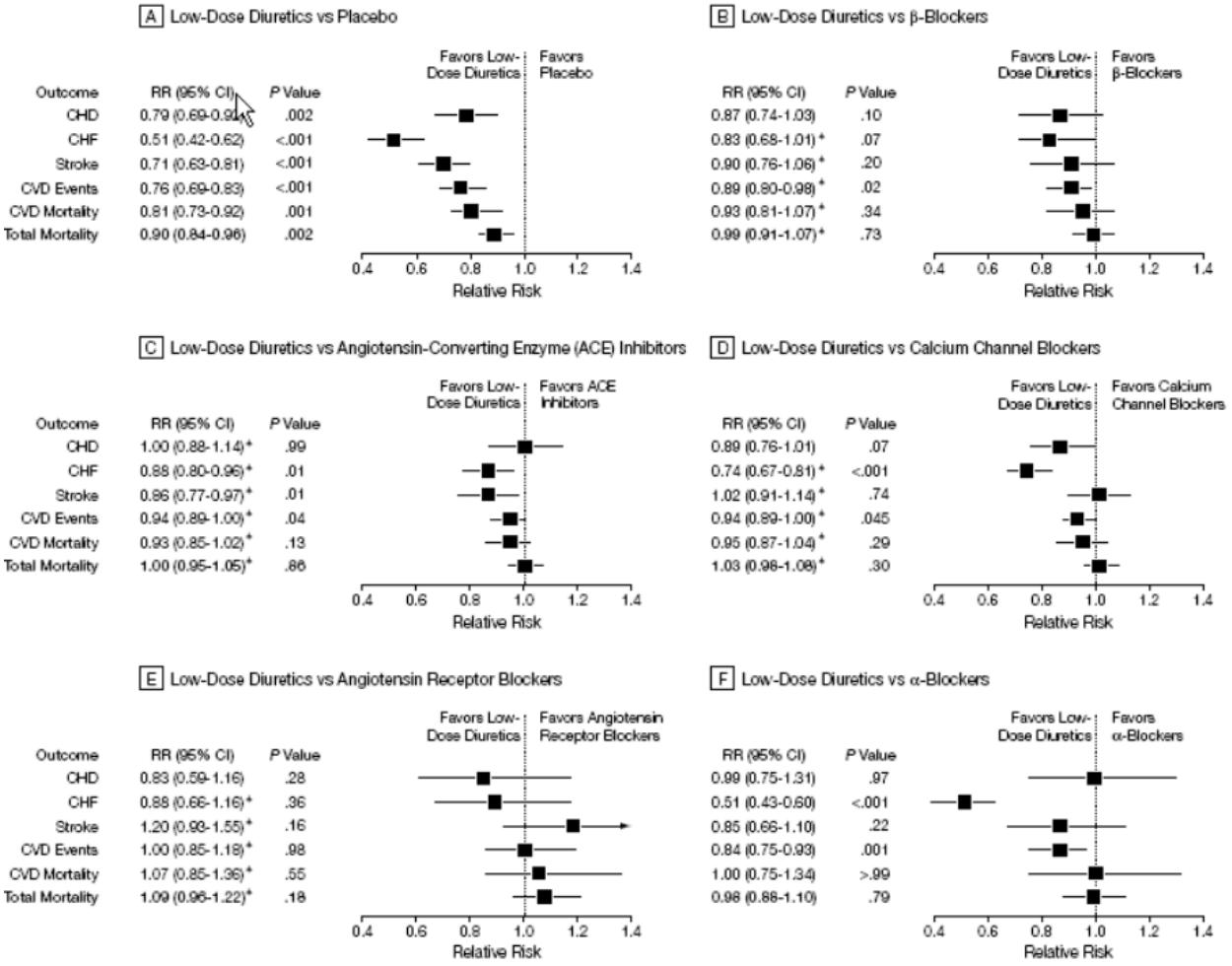


Figure 15 - Results of network meta-analysis
Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, *et al.* Health outcomes associated with various antihypertensive therapies used as first-line agents. A network meta-analysis. JAMA 2003;289(19):2534-44.

4.7 Meta-regression

4.7.1 General comments

Another method that can be used to perform indirect comparisons is meta-regression.
Meta-regression is a standard meta-analysis method that consists of modelling the effects observed in trials using covariables (25-29). Often this involves a linear regression using covariables describing the treatment evaluated (dose, medicinal products, etc.), patients included in the trials, methodological characteristics of trials, etc., as explanatory variables.
For example, meta-regression was used to investigate whether there was any relationship between the reduction in morbidity and mortality obtained by statins and the degree of reduction in LDL obtained (30).

For statistical reasons, it is better to use the logarithm of the odds ratio as a measure of the effect of treatment. The meta-regression model is therefore written:

$$\log\left(OR_i\right) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \varepsilon_i$$

where OR_i represents the odds ratio measured in the trial *i*, $x_{1,i}$ the value of covariable 1 in this trial *i*, etc. for all covariables considered. ε_i represents the random residual factor which is considered to be distributed normally with σ^2 as residual variance.
ε_i therefore represents that part of the variance of the effect between studies not explained by the covariables considered.
The coefficients β_1, β_2, … can be estimated by various adjustment methods such as least-squares, weighted least squares, etc. However, Thompson and Sharp recommend using the restricted maximum likelihood method (28).
A random model may also be used. By simultaneously estimating several indirect comparisons, meta-regression is also related to methods based on modelling a network of trials.

4.7.2 Meta-regression and fixed treatment effect

- *Statistical method*

Meta-regression is used to make indirect comparisons by coding treatments for comparison using covariables.
Let us consider a situation where 3 treatments A,B,C are being studied in a series of placebo-controlled trials. Indirect comparisons of the efficacy of these three treatments will be performed by adjustment using data provided by the series of trials following the model:

$$\log\left(OR_i\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

The treatment evaluated is coded using the two co-variables X_1 and X_2 (dummy variable method) taking treatment A as reference:

Tests	X_1	X_2
A versus placebo	0	0
B versus placebo	1	0
C versus placebo	0	1

Thus the coefficient β_1 will estimate the effect of B compared with A and coefficient β_2 that of C compared with A. Of course, any other treatment could have been used as reference.
Indeed, for treatment A (i.e. x_1=0 and x_2=0) this gives

$$\begin{aligned} Log\left(OR_A\right) &= \beta_0 + \beta_1 0 + \beta_2 0 \\ &= \beta_0 \end{aligned}$$

and for treatment B

$$\log\left(OR_B\right) = \beta_0 + \beta_1$$

Thus

$$\begin{aligned} Log\left(OR_B\right) - Log\left(OR_A\right) &= \beta_0 + \beta_1 - \beta_0 \\ &= \beta_1 \end{aligned}$$

As $Log\left(B\right) - Log\left(A\right) = Log\left(\frac{B}{A}\right)$, this gives

$$\frac{OR_B}{OR_A} = \exp\left(\beta_1\right)$$

Now $OR_B / OR_A = OR_{B/A}$, the odds ratio of B compared with A. Indeed:

$$OR_B = OR_{B/PBO} = \frac{odds_B}{odds_{PBO}}$$

Similarly

$$OR_A = \frac{odds_A}{odds_{PBO}}$$

which gives

$$\frac{OR_B}{OR_A} = \frac{odds_B}{odds_{PBO}} \frac{odds_{PBO}}{odds_A}$$
$$= \frac{odds_B}{odds_A}$$
$$= OR_{B/A}$$

The conclusion is therefore $OR_{B/A} = exp(\beta_1)$ and $OR_{C/A}=exp(\beta_2)$

Meta-regression can also be used to perform statistical tests. Thus to find out whether B is statistically different from A, it is sufficient to test the hypothesis $\beta_1=0$. Similarly, the estimated odds ratio of B/A is accompanied by its confidence interval.

- *Example of application*

As an example, meta-regression with a mixed effects model was used by Eckert and Lancon to compare duloxetine with fluoxetine and venlafaxine (31).

4.7.3 Meta-regression and random treatment effect (Bayesian method)

Nixon *et al.* proposed an implementation of meta-regression to perform mixed treatment comparisons using a Bayesian model estimated by MCMC (32). Compared with simple meta-regression, their model includes a random treatment effect and improves its performance by adjusting for covariables.

4.8 Methods for continuous endpoints

The methods proposed so far only apply to binary endpoints. Only one of these methods can be transposed directly to continuous endpoints, i.e. the method of adjusted indirect comparisons. The relative risk only has to be replaced by a measure appropriate for meta-analysis of continuous endpoints, i.e. weighted mean difference (WMD) or effect size.
In contrast, an evidence network approach cannot be directly transposed to continuous endpoints and specific methods will need to be developed.
The meta-analysis of Packer *et al.* comparing the effects of carvedilol and metoprolol on ventricular ejection fraction in heart failure is an example of the use of adjusted indirect comparisons with a continuous endpoint (33).

4.9 Comparison of the different methods

Only one study has compared the statistical performance of the different methods used to perform indirect comparisons. Glenny *et al.* used data from the International Stroke Trial (IST) to create pseudo-trials to show (Chapter 5) that the different methods produce estimates that are very similar to each other (12). The methods they compared are:

- adjusted indirect comparison using a fixed model;
- adjusted indirect comparison using a random model (DerSimonian and Laird);
- logistical regression;
- random effect meta-regression;
- naive method by comparison of arms.

The results show that with these data, all the methods produce very similar point estimates. However, for estimating standard errors (and performing statistical tests), as there are at least three levels of variability in the question posed by indirect comparisons, methods based on a random model seem to be more logical.

Table 8 attempts to compare the theoretical advantages and disadvantages of the different methods. As well as statistical aspects, it should also be noted that only adjusted indirect comparisons have been the subject of an empirical study of their validity by comparing their results to those of direct comparisons (see Chapter 6). In addition, this is the only method that can give results using relative risk. The other approaches produce odds ratios. Although the odds ratio has good statistical and mathematical properties, it is difficult to interpret. Relative risk is an intuitive approach and so seems to be more suitable for questions of clinical evaluation and meta-analyses (34).

In conclusion, in the absence of any known differences in performance between the different methods, it is difficult to recommend any individual method rather than another. However, two approaches may be pointed up:

1. adjusted indirect comparisons, because of their simplicity, transparency (the origin of the main results is clear), the possibility of expressing results as relative risk, and because this is the only standard method available for continuous endpoints;
2. Bayesian methods based on an evidence network because of the great flexibility of the model (allowing detailed and flexible modelling of data, which can be adjusted for particular cases), estimation of inconsistency, and ability to take account of multiarm trials.

Table 8 - Theoretical advantages and drawbacks of the different methods

Method	Advantages	Disadvantages
Adjusted indirect comparison	• Simple calculations • Transparency • Can be used with relative risk • Can be transposed to continuous endpoints • Empirical study of validity	• Impossible to test inconsistency • Multiple 2 x 2 comparisons (significance level inflation) if indirect comparison of more than two treatments
Meta-regression	• Adjustment for other covariables introducing heterogeneity • Does not necessarily require MCMC • Can also be estimated by MCMC	• No estimation of inconsistency • Gives odds ratios • No empirical study of validity
Methods based on evidence networks		
Lumley, 2002 (21)	• Test of incoherence • Does not require MCMC	• Gives odds ratios • No empirical study of validity
Lu and Ades, 2004 (19)	• Test of inconsistency • Takes account of multi-arm trials • Possibility of using a random model	• Difficult estimation method requiring considerable expertise (including analysis of convergence) • Bayesian approach that requires introduction of prior information • Gives odds ratios • No empirical study of validity
Caldwell *et al,* 2005 (20)	• Test of inconsistency • Takes account of multi-arm trials • Possibility of using a random model	• Difficult estimation method requiring considerable expertise (including analysis of convergence) • Bayesian approach that requires introduction of prior information • Gives odds ratios • No empirical study of validity

MCMC: Markov chain Monte Carlo

5. Current use of indirect comparisons

5.1 Bibliographical data

A systematic review of the use of indirect comparison methods was published in 2005 by Glenny *et al.* (12). In this study, systematic reviews published before 1999 including indirect comparisons were searched for using the Database of Abstracts of Reviews of Effects (DARE) and the Cochrane Database of Systematic Reviews (CDSR) .

Thirty-six studies involving indirect comparisons were found; thirteen of them involved simultaneously direct and indirect comparisons. In eleven of the thirty-six cases (31%), the indirect comparisons were performed improperly by naive comparison of active arms. In the other twenty-five cases (69%), the adjusted indirect comparison method was applied.

5.2 Some recent publications

As the study by Glenny *et al.* is now rather old (search stopped in 1999), we carried out a non-exhaustive additional search of more recent studies (12) (see Section 3). This search was not carried out to quantify how often these methods are used, but simply to provide an overview of current methods and the fields in which they are used.

The literature search was complicated by the fact that there is no unique term to represent the concept of indirect comparison. In addition, indirect comparisons are used within meta-analyses without being highlighted in the title or in the abstract, which makes it even more difficult to identify them. For this study, the literature search used the following terms:

- *Indirect comparison*;
- *Mixed treatment comparison*;
- *Multiple treatment comparison*.

The result of this overview is shown in Table 9. Naive methods are still found. Most of the studies used adjusted indirect comparison. Bayesian methods on networks of comparisons are increasingly found.

Table 9 - Recent examples of studies including an indirect comparison approach

Publication title (ref.)	Year	Method
Gabapentin versus tricyclic antidepressants for diabetic neuropathy and post-herpetic neuralgia: discrepancies between direct and indirect meta-analyses of randomized controlled trials (35)	2009	Adjusted indirect comparison
Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis (36)	2009	Network meta-analysis (Bayesian method)
Carbapenems versus other beta-lactams in the treatment of hospitalised patients with infection: a mixed treatment comparison (37)	2009	Mixed Treatment Comparison
An indirect comparison of the efficacy and safety of factor Xa inhibitors with thrombin inhibitors in preventing various thromboembolism after hip or knee surgery (38)	2008	Mixed Treatment Comparison
Meta-analysis of the combination of TNF inhibitors plus MTX compared to MTX monotherapy, and the adjusted indirect comparison of TNF inhibitors in patients suffering from active rheumatoid arthritis (39)	2008	Adjusted indirect comparison
Multiple-treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer (40)	2008	Network meta-analysis (Bayesian method)
Direct and indirect comparison meta-analysis demonstrates the superiority of sirolimus- versus paclitaxel-eluting stents across 5854 patients (41)	2007	Adjusted indirect comparison
Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis (18)	2007	Bayesian method

Publication title (ref.)	Year	Method
Need for additional calcium to reduce the risk of hip fracture with vitamin D supplementation: evidence from a comparative meta-analysis of randomized controlled trials (42)	2007	Adjusted indirect comparison
Magnetic resonance colonography versus computed tomography colonography for the diagnosis of colorectal cancer: an indirect comparison (43)	2007	Adjusted indirect comparison of diagnostics odds ratios (DOR)
What is the optimal pharmacological prophylaxis for the prevention of deep-vein thrombosis and pulmonary embolism in patients with acute ischemic stroke? (44)	2007	Naive comparison of point estimates
Adjusted indirect comparison of celecoxib versus rofecoxib on cardiovascular risk (45)	2006	Adjusted indirect comparison
Duloxetine compared with fluoxetine and venlafaxine: use of meta-regression analysis for indirect comparisons (31)	2006	Meta-regression
Incorporating direct and indirect evidence using Bayesian methods: an applied case study in ovarian cancer (46)	2006	Mixed Treatment Comparison
Are statins created equal? Evidence from randomized trials of pravastatin, simvastatin, and atorvastatin for cardiovascular disease prevention (47)	2006	Adjusted indirect comparison
Efficacy of drug eluting stents in patients with and without diabetes mellitus: indirect comparison of controlled trials (48)	2006	Meta-regression and adjusted indirect comparisons
Efficacy of PDE-5-inhibitors for erectile dysfunction. A comparative meta-analysis of fixed-dose regimen randomized controlled trials administering the International Index of Erectile Function in broad-spectrum populations (49)	2006	Adjusted indirect comparison (with Bonferroni)
Investigating heterogeneity in studies of resting energy expenditure in persons with HIV/AIDS: a meta-analysis (50)	2005	Bayesian method
Adjusted indirect comparison of intracoronary drug-eluting stents: evidence from a meta-analysis of randomized bare-metal-stent-controlled trials (51)	2005	Adjusted indirect comparison
Ximelagatran compared with warfarin for the prevention of systemic embolism and stroke. An imputed placebo analysis (52)	2005	Inappropriate method
Indirect comparison of interventions using published randomised trials: systematic review of PDE-5 inhibitors for erectile dysfunction (53)	2005	Adjusted indirect comparison
A hierarchical Bayesian meta-analysis of randomised clinical trials of drug-eluting stents (54)	2004	Bayesian method
Clinical efficacy of antiretroviral combination therapy based on protease inhibitors or non-nucleoside analogue reverse transcriptase inhibitors: indirect comparison of controlled trials (55)	2004	Adjusted indirect comparison
Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation (56)	2004	Adjusted indirect comparison
Ultrasonic locating devices for central venous cannulation: meta-analysis (57)	2003	Unidentified method
Indirect comparison meta-analysis of aspirin therapy after coronary surgery (58)	2003	Adjusted indirect comparison
Effectiveness of nifedipine versus atosiban for tocolysis in preterm labour: a meta-analysis with an indirect comparison of randomised trials (59)	2003	Adjusted indirect comparison
Treatment of open fractures of the shaft of the tibia. A systematic overview and meta-analysis (60)	2001	Adjusted indirect comparison

Publication title (ref.)	Year	Method
Comparative effects of carvedilol and metoprolol on left ventricular ejection fraction in heart failure: results of a meta-analysis (33)	2001	Adjusted indirect comparison on continuous endpoints

5.3 Use in health technology assessments and guidelines

An indirect comparison approach was used in some recent health technology assessments and guidelines (list not exhaustive):

- Manifestations and management of chronic insomnia in adults, Agency for Healthcare Research and Quality, 2005 (61)
- Metal-on-metal total hip resurfacing, TEC, Blue Cross Blue Shield Association, 2007 (62)
- Off-label uses of bevacizumab: renal cell carcinoma and other miscellaneous non-colorectal cancer indications, TEC, Blue Cross BlueShield Association, 2006 (63)
- Botulinum toxin type A therapy for cervical dystonia, Cochrane Review, 2005 (64)
- Topical antibiotics without steroids for chronically discharging ears with underlying eardrum perforations, Cochrane Review, 2005 (65)
- Docetaxel for the adjuvant treatment of early node-positive breast cancer, National Institute for Health and Clinical Excellence, 2006 (66)
- Pemetrexed for the treatment of non-small-cell lung cancer, National Institute for Health and Clinical Excellence, 2007 (67)
- Adalimumab for the treatment of psoriatic arthritis, National Institute for Health and Clinical Excellence, 2007 (68)
- Rituximab for the treatment of rheumatoid arthritis, National Institute for Health and Clinical Excellence, 2007 (69)
- Comparative effectiveness and safety of oral diabetes medications for adults with type 2 diabetes, Agency for Healthcare Research and Quality, 2007 (70)
- Infliximab for the treatment of adults with psoriasis, National Institute for Health and Clinical Excellence, 2008 (71)
- Varenicline for smoking cessation, National Institute for Health and Clinical Excellence, 2007 (72)

6. Validity of indirect comparisons

6.1 How valid are indirect comparisons?

The classical reference approach for comparing two active treatments is direct comparison. For example, ICH E10 (choice of control group in clinical trials) states (Section 2.1.7.4): "Placebo-controlled trials lacking an active control give little useful information about comparative effectiveness, information that is of interest and importance in many circumstances. Such information cannot reliably be obtained from cross-study comparisons, as the conditions of the studies may have been quite different."
Starting from this assumption, for indirect comparisons to be regarded as valid they should produce the same results in a given situation as those obtained from direct comparisons.
Therefore, the validity study of indirect comparisons implies the study of differences between the results obtained by both methods. This approach means having fields in which both methods are applicable, i.e. fields in which there are direct comparisons and materials making it possible to carry out indirect comparisons.

6.2 Validity studies

6.2.1 Studies of Song *et al.*, 2003 and of Glenny *et al.*, 2005

The most complete study addressing the question of the validity of indirect comparisons is that of Song *et al.* published in the BMJ in 2003 (14). The study was reported in more detail in the NHS R&D HTA Programme document coordinated by Glenny *et al.* (12).

In this study, indirect comparisons were performed with the adjusted indirect comparison method. The measure of discrepancy used was the difference $\Delta = T_{direct} - T_{indirect}$ where T_{direct} and $T_{indirect}$ represent the direct and indirect estimates respectively of the difference in effect between the two treatments. Fields in which direct comparisons exist and in which it is possible to make indirect comparisons were looked for. T_{direct} and $T_{indirect}$ were obtained by meta-analyses using a random model. For continuous endpoints, estimates of the treatment effect concerned the difference in means while for binary endpoints, the logarithm of relative of risk was used.
It is possible to calculate a 95% confidence interval of the difference Δ. If the direct and indirect estimates were identical, the difference Δ was nil.
Figure 16 shows the degrees of discrepancy and their confidence interval observed for the forty-four meta-analyses studied in different clinical fields.
The confidence interval of the difference Δ makes it possible to incorporate into comparison of results from both methods the statistical uncertainties existing at their level. It was essential to take account of these uncertainties since the reference (T_{direct}) was not known with precision but was estimated with a margin of error. So this confidence interval represents the range of values reasonably compatible with the true difference between the two methods taking account of the possible margin of error for T_{direct} and $T_{indirect}$. When this interval does not include zero, it may be concluded that there is a real difference between the two estimates T_{direct} and $T_{indirect}$, even despite the uncertainties weighing on T_{direct} and $T_{indirect}$. Conversely, if the interval contains zero, the difference observed may lead to errors in the estimates T_{direct} and $T_{indirect}$ and it is therefore not possible to conclude that there is any real difference in results between the two estimates.

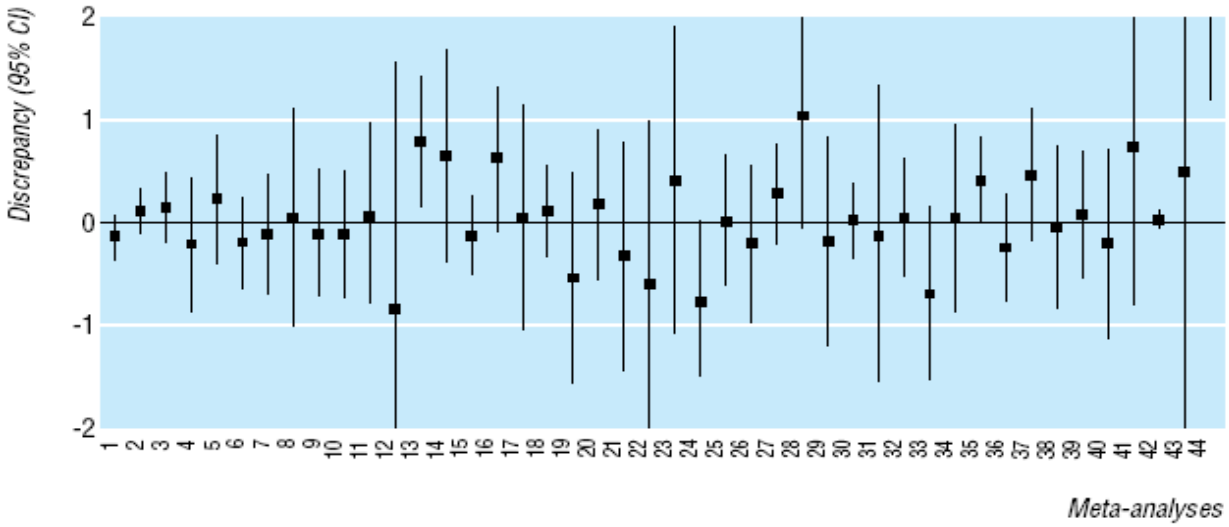


Figure 16 - Discrepancy between the results of direct and indirect comparisons for the forty-four meta-analyses studied by Song *et al.*, 2003 (14)
Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 2003;326(7387):472. Copyright 2003, with permission from BMJ Publishing Group Ltd.

In this study, only three out of the forty-four cases found a statistically significant disagreement between the estimate obtained from direct comparison and that from indirect comparison. Two of these cases of disagreement concern treatment with H2 antihistamines or proton pump inhibitors. The remaining case concerns paracetamol in combination with codeine in the treatment of post-surgical pain (12,73); complete discrepancy was found between the two estimates (Table 10).

Table 10 - Explanation of discrepancies observed between direct and indirect comparisons in the case of a combination of paracetamol and codeine for the treatment of surgical pain in the study by Glenny *et al.*, 2005 (12)

	Pain intensity score **Difference in means between groups (95% CI)**	
	Direct comparison	**Adjusted indirect comparison**
All trials	6.97 (3.56-10.37) N=13	-1.16 (-6.95-4.64) N=12 and 31
Trials of dental surgery only	7.07 (3.37-10.78) N=11	-1.40 (-8.27-5.46) N=7 and 15
Trials using 600-650 mg of paracetamol and 60 mg of codeine	7.28 (3.69-10.87) N=10	5.72 (-5.37-16.81) N=2 and 12

However, these two results were obtained by combining trials with considerable heterogeneity in terms of type of surgery and dose administered (between 300 mg and 1 000 mg for paracetamol and between 30 and 60 mg for codeine).
The discrepancy does not seem to be explained by type of surgery, as it was not eliminated by restricting the analysis to trials of post-dental surgery pain (the most numerous).
In contrast, direct and indirect comparison do produce identical results when the analysis includes trials using treatments at the same dose (600-650 mg paracetamol and 60 mg of codeine). The lesson from this example is fairly trivial. It is simply the fact that in a meta-analysis (whether or not intended for indirect comparisons) it is necessary to group together trials that are similar with regard to the question posed, and to analyse sensitivity. Here, it is possible to suggest a relationship between dose and effect to explain the discrepancies observed, even though other unidentified factors may also have contributed to the result.

6.2.2 Study by Bucher *et al.,* 1997

The first study of comparison was carried out in 1997 by Bucher *et al.* (74). The study concerned prevention of *Pneumocystis Carinii* pneumonia in patients with HIV infection and comparison of the combination trimethropim-sulfamethoxazole with the combination dapsone-pyrimethamine.
Eight trials performing a direct comparison were available, giving an odds ratio of 0.64 in the meta-analysis. Estimation using adjusted indirect comparison based on the results of fourteen trials gave an odds ratio of 0.37. A comparison test of these two estimates gave p=0.11.
In this example, the two approaches did not lead to the same estimates even though they remain compatible in view of their degrees of precision (p for heterogeneity = 0.11).
The authors compared the characteristics of the studies without finding any real explanation for this discrepancy.

6.2.3 Study of Song *et al.,* 2000

Prior to their 2003 study, Song *et al.* (75) had carried out a first comparison of the results obtained from direct and indirect comparisons for antibiotic prophylaxis in colorectal surgery. A high level of discrepancy was observed. The authors attributed this discrepancy to chance because of the small number of trials and their small populations. In fact, only three multiarm trials were involved in the calculations, simultaneously providing the data required for the indirect comparisons and for the direct comparisons. The fact that different arms within a single trial were not independent was not taken into consideration.
As it only studied a single subject (corresponding to only one clinical question) and as only a small number of trials were involved, the study cannot be considered an empirical study of the validity of indirect comparisons.

6.3 Is direct comparison still the gold standard?

Direct comparison trials also called head-to-head trials, are currently regarded as the gold standard approach for comparing the efficacy and safety of two active treatments, even though in practice this type of study is often lacking (see Section 2.2). The approach has a number of advantages. It answers the question of the relative efficacy of two treatments with all the methodological protections of the double-blind randomised controlled trial. The study population in which the result was obtained is clearly defined and known.
However, in practice, there is a number of limitations related to direct comparisons.

6.3.1 Limitations of direct comparison trials

The first limitation is that this approach is not frequently used (see Section 2.2). In many clinical fields there are no direct comparison trials between active treatments, particularly once Marketing Authorisation has been granted, so there are no answers to the question of the hierarchy of efficacy and/or safety of the competing treatments concerned.
Often, direct comparison trials are carried out after the registration dossier is produced, and the methodological and quality requirements for these trials can be lower than for trials intended for registration (76).

Direct comparison trials between active treatments are often found to be of a lower level of evidence than placebo-controlled trials, particularly with regard to blinding. There are technical, pharmaceutical and analytical problems with the masking of active treatments which have often already been placed on the market. As an example, encapsulation can modify the kinetics of medicinal products and lead to changes in efficacy, so biasing the results of head-to-head trials (77,78).

The double placebo method raises specific problems with regulatory and commercial issues. When the treatments compared are not medicinal products, it is even more difficult to achieve a double-blind. The trials are generally not carried out double-blind, with a few exceptions (versus sham, for example).

Another disadvantage in practice is that these trials are usually carried out at a late stage, after a marketing authorisation has been obtained and the product placed on the market, which leaves the initial evaluation without any direct comparator to define the place of a new treatment in the hierarchy of clinical resources in a specific disease.

Finally, there is a constant issue with assay sensitivity against the active treatment. This means the ability of an assay to detect any differences. This ability depends on the best use of treatments (dose, regimen), patients included, statistical power, performance of the endpoints, etc. At the regulatory level, it is strongly recommended (ICH E10, Section 1.5) that a placebo group be included in trials comparing two active treatments to demonstrate that assays are able to identify differences when they exist. As an example, several trials have compared a new triptan to a reference triptan in the treatment of acute migraine attacks (79).

It is increasingly common for trials comparing two active treatments to be non-inferiority trials, which raise very specific problems. In this type of study, the conclusion of non-inferiority is made at the cost of a loss of potential efficacy. So this type of trial is not really able to answer the question of the hierarchy of treatments.

Table 11 - Advantages and limitations of direct comparison trials between treatments

Advantages	Limitations
• Direct measure of the question posed • High potential degree of protection against bias • No hypothesis on the consistency of effects of A and B	• Rarely carried out • Rarely double-blind • Often carried out with lower quality standards than those for placebo-controlled trials • Carried out late in the development of new treatments • Internal validation rare

Table 12 - Advantages and limitations of direct and indirect comparisons relating to the gathering and the quality of data analysed

	Advantages and limitations	
	Direct comparison	indirect comparison
Time to obtaining data	• Late	• Early (as soon as the first trials are available)
Methodological quality of trials	• Frequent failure of blinding • Quality variable but nothing to exclude highest methodological quality	• Uses results obtained with double-blind method • Uses results produced to highest quality requirements
Exhaustive nature of comparisons with all competing treatments in the field	• Rare • Limited in some fields • Impossibility to use data globally	• Consistent with methods based on a network of comparisons
Publication bias	• Unlikely because of the size of populations required	• Can be avoided by including all trials carried out • Contribution of clinical trials registry, likely to improve the relevance of indirect comparisons

6.3.2 Case study

In Alzheimer's disease a direct comparison between donepezil and galantamine has been documented in only two trials, one published by Jones *et al.* (80) and the other by Wilcock *et al.* (81). Table 13 describes these two trials and their results. Each of the two trials was funded by a manufacturer of one or other of the drugs. The trial by Jones *et al.* comparing donepezil with

galantamine was sponsored by the manufacturer of donepezil while the trial of Wilcock *et al.* comparing galantamine with donepezil was funded by the manufacturer of galantamine.
These two trials gave contradictory results, each time concluding in favour of the drug produced by the trial sponsor.
Obviously, this example cannot be generalised but it does demonstrate that a head-to-head trial may not be the solution to the problem of comparing two treatments. In particular, the discrepancies in the results may be explained by the fact that these were open trials.

In a recent article, Song *et al.* proposed three further case studies in which indirect comparisons were found to be more conservative than direct comparisons, again raising the question of the validity of these studies (82).

Table 13 - Description and results of two direct comparison trials of donepezil (DON) versus galantamine (GAL)

	Jones *et al.*, 2004 (80)	Wilcock ,*et al.* 2003 (81)
Company	Manufacturer of donepezil (Eisai Inc., Teaneck, NJ, United States, Pfizer Global Pharmaceuticals, Pfizer Inc., NY, United States)	Manufacturer of galantamine (Janssen-Cilag, Johnson & Johnson Pharmaceutical Research, Shire Pharmaceuticals)
Patients	Mild to moderate Alzheimer's disease	Alzheimer's disease
Comparison	donepezil (up to 10 mg/day) or galantamine (up to 24 mg/day)	galantamine (24 mg/day) or donepezil (10 mg/day)
Duration	12 weeks	53 weeks
Population	120	182
Primary/secondary endpoints	Physicians' and caregivers' satisfaction with treatment/ease of use in daily practice / ADAS-cog, MMSE, DAD-ADL	BrADL, MMSE, ADAS-cog/11, NPI, Screen for Caregiver Burden
Blind	Open	Open, rater-blinded
Trial conclusion	Physician and caregiver ease of use/satisfaction scores, and assessments of cognition and ADL, showed significant benefits for donepezil compared with galantamine	Significant advantages were found in the treatment response to galantamine (versus donepezil) on cognition as measured by response rates on the MMSE and ADAS-cog/11

6.4 Outlook

A summary meta-analysis of empirical studies investigating the validity of indirect comparisons is planned as part of the Cochrane Collaboration by Song *et al.*[12]. For the moment, it is only at the protocol stage. Only one empirical study of this type is currently available.

[12] Song F, Altman DG, Glenny A, Eastwood AJ, Deeks JJ. Adjusted indirect comparison for estimating relative effects of competing healthcare interventions (Protocol). Cochrane Database of Systematic Reviews 2007, Issue 2.

7. Position of regulatory bodies

European Medicines Agency: no guidelines on indirect comparisons were found on the EMEA website (search carried out on 01 October 2007).

Federal Drug Administration: no guidelines on indirect comparisons were found on the FDA website (search carried out on 01 October 2007).

National Institute for Health and Clinical Excellence: the NICE procedure documents mention indirect comparisons. In the context of producing single technology appraisals (STA), a document aimed at companies, "Specification for manufacturer/sponsor submission of evidence" (83), mentions indirect comparisons in the following terms:
"The Institute has a strong preference for evidence from 'head-to-head' randomised controlled trials (RCTs) that directly compare the technology and the appropriate comparator(s). Wherever such evidence is available, and includes relevant outcome evidence, this is preferred over evidence obtained from other study designs. Where no head–to-head RCTs are available, consideration will be given to indirect comparisons, subject to careful and fully described analysis and interpretation. "

The same text appears in the "Guide to the methods of technology appraisal" (84).

Haute Autorité de Santé: no guidelines on indirect comparisons were found on the *HAS* website (search carried out on 01 October 2007).

Canadian Agency for Drugs and Technologies in Health: methodology summary report published in March 2009 (85).

A recent article discusses the potential benefit of using indirect comparison and network meta-analysis methods in this context (4).

8. Interpretation of results

There are two types of potential errors when interpreting the results of indirect comparisons, mainly those derived from networks of comparisons:

1. drawing conclusions of equivalent efficacy for two treatments when there is no statistically significant difference
2. and within an indirect comparison, establishing an incorrect hierarchy by naive comparison of point estimates.

For example, based on the results obtained from a meta-analysis of the effect of antihypertensives on onset of diabetes (18), it is clearly not legitimate to conclude that beta-blockers are the same as diuretics. Similarly, the conclusion that angiotensin receptor blockers (ARB) lead to fewer new cases of diabetes than ACE inhibitors or calcium channel blockers (CCB) is based on a naive comparison of point estimates.
The ideal way of avoiding this difficulty would be to take the treatment thought to be the best tolerated as reference and to show that all the others are statistically significantly inferior to it.

9. Critical review guide

The aim of the critical review guide offered in this chapter is to check whether the indirect comparison matches the ideal situation as described in the section below and may be acceptable. If not, the unverified points each represent a reservation about the validity of the available results. In the light of a recent empirical study, it appears that many indirect comparison studies contain major methodological problems, which makes it essential to carry out an in-depth critical review of these studies (86).

9.1 Ideal situation

To be considered acceptable, a result based on indirect comparison should satisfy the following conditions:

- the indirect comparison was carried out using a suitable statistical method, i.e. adjusted indirect comparison, mixed linear model, Bayesian methods, meta-regression, etc. (see Chapter 1);
- a satisfactory meta-analysis was performed to collect the trials contributing to indirect comparisons (exhaustive search of published and unpublished trials, rigorous selection based on methodological quality of trials and a thorough analysis of heterogeneity to document the hypothesis of stability of effects, see below);
- trials used in indirect comparisons were similar in terms of treatment (dose, dosage regimen), clinical situation (concomitant treatments including interventions other than drugs), disease (definition, diagnostic criteria, criteria of severity and progression of disease), timing (same treatment period and same secular shift). Subgroup analyses in the trials are reassuring with regard to lack of heterogeneity of treatment effect in relation to these covariables[13];
- if a number of paths are available in the network of trials, the data allow to test the inconsistency in a relatively satisfactory way and no inconsistency has been found;
- the results of existing direct comparisons do not differ from the results of a mixed treatment comparison obtained by combining direct and indirect comparisons which will be highlighted;
- the results are being interpreted correctly. In particular, no conclusion of equivalence has been made based on observation of absence of difference.

9.2 Critical review guide

This critical review guide for results obtained by indirect comparison consists of six main sections designed to check the following:

1. acceptability of the approach used;
2. search strategy and selection process for data contributing to the indirect comparison calculations;
3. clinical homogeneity of trials and stability of effects;
4. consistency of estimates;
5. degree of concordance of the result with that of existing direct comparisons;
6. correct interpretation of results in the proposed conclusions.

[13] Absence of evidence of major variation in effect of treatment in relation to these covariables.

9.2.1 Analysis of the approach used

First it should be verified that the method used is suitable for indirect comparison. This implies the exclusion of inappropriate methods. All other methods (see above) are acceptable.

i. Reject a naive comparison of point estimates derived from different controlled trials

An approach comparing point estimates with each other without taking account of the confidence intervals should be rejected (see Section 4.1.1). An example of a result of this type is given in the case study in Section 9.3.1.

ii. Reject a naive comparison of the active arms of different controlled trials

An approach comparing the different active arms of trials without taking account of the control arms, breaking the randomisation, should be rejected (see Section 4.1.2).
An example is given in the case study in Section 9.3.2.

9.2.2 Analysis of the search and selection of trials for inclusion

This stage examines the process of how the list of trials was established for inclusion in the meta-analyses required for indirect comparison calculations: were trials selected to give the desired indirect comparison results or were they all the trials free of bias carried out in the field?

i. Ensure that the search for trials was exhaustive

Was a rigorous, exhaustive and reproducible search process for meta-analysis trials followed?
Were unpublished trials searched for? Like classical meta-analysis, indirect comparisons are sensitive to problems of publication bias.

ii. Eliminate opportune selection of trials

This means eliminating the possibility that trial selection was influenced by the anticipated result. At this level, critical analysis is identical to that of a meta-analysis. Strong reservations should be made if there are no selection criteria for trials or if arbitrary selection criteria were used.
At this level, an indirect comparison study should have the same properties as a classical meta-analysis. Analysing the list of trials excluded and the reasons for their exclusion is useful when analysing the selection of trials with a methodological quality which is a sufficient guarantee of absence of bias.

iii. Ensure that all basic comparisons that could be used were considered during trial selection

One point related to the search for inconsistency (see below) is to ensure that all basic comparisons that could be used to make the indirect comparison being searched for, were indeed included. For an indirect comparison of A versus B, there could be many usable basic comparisons: A versus PBO and B versus PBO, as well as A versus D and B versus D, etc. The greater the diversity of the basic comparisons, the more powerful will be the test of hypothetical inconsistency.

9.2.3 Analysis of clinical homogeneity of trials and stability of effects

This stage checks that the trials searched for and included were all trying to answer the same clinical question with regard to type of patients studied and treatments compared. This analysis will also verify the conditions of acceptability of the hypothesis of stability of effects through trials providing the basic comparisons.

iv. Homogeneity of treatments studied in the trials

Did the trials study the same treatment modalities (in terms of dose and dosage regimen)? The example of combinations of codeine and paracetamol (see Section 6.2.1) is a perfect illustration of this point.

v. Absence of interaction factor or homogeneity of populations with regard to interaction variables

In order to verify the plausibility of the hypothesis of stability of effects which is essential to the validity of indirect comparisons (see Section 4.2.2), it has to be verified that the subgroup analyses performed in the trials did not reveal any interaction variables modifying the effects of treatments. If such variables do exist, the validity of indirect comparisons may be compromised if it is found that the populations of trials of A versus control and of B versus control were not similar in terms of these covariables.
The case study in Section 9.3.7 gives an example of a situation in which an interaction problem was raised.

vi. Absence of statistical heterogeneity between studies of the same comparison

The presence of heterogeneity in the clinically significant statistical results between studies of A versus control or of B versus control is likely to compromise the hypothesis of stability of effects. If the meta-analysis finds differences in the results of trials performing the same comparison, this suggests a variability of effects between trials, and therefore the existence of one or more interaction factors.
In the absence of heterogeneity, it has not been possible to demonstrate any change in effect between trials. If some trials of A and of B are similar in terms of contexts and types of patient, this suggests that the hypothesis of stability of effects may reasonably be accepted.

9.2.4 Analysis of the hypothesis of consistency

The result of an indirect comparison is clinically meaningful if the result is found to be the same irrespective of the chain of comparisons used to obtain it (see Section 4.4), in other words, if the hypothesis of consistency is verified. An indirect comparison result will be all the stronger if absence of inconsistency can be proved with power. Reservations should be raised whenever any (clinically relevant) inconsistency has been detected or if it has not been possible to look for it.
This point should be analysed for each indirect comparison present in a study as it concerns the results, rather than just the method.

vii. Has inconsistency been tested for?

Was testing for inconsistency planned? Did the method used make it possible to test for inconsistency?

viii. Did the test for inconsistency have a minimum efficacy?

The ability to demonstrate a lack of consistency will depend on a number of trials carrying out the same comparisons and on the diversity of the pivotal comparators (indirect comparison paths). If there is only one path (A versus PBO and B versus PBO) inconsistency cannot be tested; the only thing that can be tested is the heterogeneity of estimates of both types of comparison.

9.2.5 Analysis of the statistical method used

This is the most technical point of the review guide. The aim is to verify that the statistical method has been applied correctly. This point is included in the guide because some indirect comparison methods are difficult to apply, particularly those using Markov chain Monte Carlo (MCMC) and Gibbs sampling methods. In practice, it is often necessary to consult experts in statistics to decide on the technical acceptability of this type of result.

For methods using estimation by MCMC and Gibbs sampling:

Was a convergence diagnosis performed? Ideally, this diagnosis involves at least one graphical analysis of the history of the two chains. The use of diagnostic tools such as CODA and BOA is strongly recommended.

Are the results obtained independent of the chosen priors? A sensitivity study of the results showing that they are independent of the choice of prior is essential.
Were the chosen priors really uninformative? In general, it is recommended that a variance of 100 should be used at least for normal distributions and that a gamma distribution (0.0001, 0.0001) should be used for precisions. In particular, situations in which an informative prior is required to obtain convergence should be rejected.
An example of a satisfactory use of these methods is reported in the case study in Section 9.3.5.

9.2.6 Were the results of indirect comparisons reinforced by the availability of direct comparisons?

ix. Were there any direct comparisons?

Was there a search for direct comparison trials?

x. Is there concordance between the direct comparison and indirect comparison estimates?

Are the direct comparison estimates compatible with the indirect or mixed comparison estimates? This concordance is evaluated using a heterogeneity test for both types of estimate.

xi. Does the main result include direct comparisons?

If direct comparison trials were available, does the main result incorporate them (i.e. a mixed treatment comparison)? In this case, the best possible estimate is one which incorporates all the available information.

9.2.7 Interpretation of results

xii. No conclusions of equivalence were based on a non-significant result

It is improper to conclude that there is no difference in efficacy on the basis of non-significant results from an indirect comparison. The problem is exactly the same as that found in an inconclusive clinical trial.

xiii. No hierarchy was proposed from naive comparison of point estimates

There is a risk of constructing a hierarchy of competing treatments by a process of naive comparison of point estimates in indirect comparisons based on networks (see Section 8). In this case, the conclusion raises the problems mentioned in Section 4.1.1, even if there is an underlying use of an appropriate indirect comparison method.

9.3 Case studies

The aim of these case studies is to illustrate some points in the critical review guide by explaining the concepts by means of examples.

9.3.1 Example of naive comparison of point estimates

An example of inappropriate indirect comparison by simple comparison of point estimates is provided by a meta-analysis involving prevention of deep vein thrombosis during the acute phase of stroke (44). The conclusion of the article was "Indirect comparison of low and high doses of UFH and LMWH suggests that low-dose LMWH have the best benefit/risk ratio in patients with acute ischemic stroke by decreasing the risk of both DVT and pulmonary embolism, without a clear increase in ICH or ECH."

The study compared high and low doses of heparin (unfractionated or low molecular weight heparin) in preventing deep vein thrombosis in the early stage of stroke. The meta-analysis was performed using trial subgroups as shown in Table 14.

Table 14 - Subgroup analysis according to anticoagulant dose, after Kamphuisen and Agnelli, 2007 (44)

	Treatment n/N	Control n/N	OR (95% CI)
High doses of heparin	2/816	26/754	0.10 (0.03-0.31)
Low doses of heparin	60/482	175/432	0.21 (0.14-0.30)

n: number of events; N: number of patients

9.3.2 Example of naive comparison of active arms of controlled trials

A meta-analysis published in 2005 in Ophthalmology (87) gives an example of inappropriate indirect comparison. The aim of the meta-analysis was to compare the efficacy against intraocular pressure (IOP) of different treatments currently used in glaucoma.
The results show the mean before-after change in intraocular pressure observed in the different arms of the trials, each treatment having been analysed without taking account of the control arm. The results were then used to rank the treatments according to these change values, leading to the conclusion "This meta-analysis suggests that bimatoprost, travoprost, latanoprost, and timolol are the most effective intraocular pressure-reducing agents in POAG and OH patients" (87).

This approach is inappropriate as it breaks the randomisation and consists of presenting observational values without any consideration of confounding factors by means of control groups (see Section 4.1.2). One would expect to find that the table of results includes estimates of efficacy by comparison of treatments rather than estimates attached to a single treatment (with no mention of the comparator).

9.3.3 Example of adjusted indirect comparison

A meta-analysis carried out by Zhou *et al.* on statins in prevention of cardiovascular disease used adjusted indirect comparisons (47). Table 15 shows, for each endpoint, the results of basic comparisons (simvastatin versus pravastatin and atorvastatin versus simvastatin) and the result of the adjusted indirect comparison of atorvastatin versus pravastatin.

Table 15 - Example of adjusted indirect comparisons in the meta-analysis of Zhou *et al.,* 2006 (47)

	Relative risk (95% CI)	p
Major coronary events		
Simvastatin versus pravastatin	0.93 (0.84-1.03)	0.18
Atorvastatin versus simvastatin	0.84 (0.66-1.08)	0.18
Atorvastatin versus pravastatin	0.79 (0.61-1.02)	0.06
Major cerebrovascular events		
Simvastatin versus pravastatin	0.87 (0.71-1.07)	0.18
Atorvastatin versus simvastatin	0.90 (0.68-1.20)	0.47
Atorvastatin versus pravastatin	0.78 (0.57-1.07)	0.12

9.3.4 Example of mixed treatment comparison

An example of the use of mixed treatment comparisons (MTC) is given by the comparison of three types of second-line chemotherapy in ovarian cancer, i.e. topotecan, paclitaxel and pegylated doxorubicin liposomes (PDL) (46).

Three trials are available documenting the three possible 2 x 2 comparisons for these three treatments:

	PDL	Topotecan	Paclitaxel
Trial A	X	X	
Trial B		X	X
Trial C	X		X

Results of the three trials for total mortality show inconsistency as topotecan appears to be inferior to paclitaxel while being superior to PDL, while paclitaxel appears to be inferior to PDL:

		Overall survival Hazard ratio (95% CI)
Trial A	Topotecan versus paclitaxel	0.914 (0.681-1.226)
Trial B	Topotecan versus PDL	1.216 (1-1.478)
Trial C	Paclitaxel versus PDL	0.931 (0.702-1.234)

Carrying out a mixed treatment comparison makes it possible to reconcile these results by no longer showing any statistically significant difference among the three treatments, suggesting the hypothesis that the initial inconsistency was caused by chance:

	MTC - Global survival Hazard ratio (95% confidence interval)
Topotecan versus paclitaxel	1.06 (0.85-1.33)
Topotecan versus PDL	1.14 (0.96-1.36)
Paclitaxel versus PDL	1.07 (0.86-1.34)

9.3.5 Example of the satisfactory use of an estimation method based on a Monte Carlo Markov chain (MCMC) method

The meta-analysis of Babapulle *et al.* concerning coated stents is an example of a satisfactory use of an MCMC method. Their description of the method shows that the priors used were sufficiently uninformative, and they state that the result was not sensitive to the priors used. Only a convergence diagnosis is missing (54).

> "Substantive prior knowledge can thereby be included into any Bayesian analysis by choice of initial (pre-data) distribution. However, because we incorporated all relevant past studies, we wanted our final (posterior) distribution to reflect the information in our dataset only and not to be influenced by our choice of initial (prior) distribution. Therefore, low-information prior distributions were used throughout, so that the data from the trials dominated the final inferences. In particular, we used normal (mean=0, variance=100) prior distributions for all means and gamma (0·0001, 0·0001) prior distributions for all precisions (which is defined as the reciprocal of the variance). Sensitivity analyses with different choices of low-information prior distributions showed robustness to this choice. In particular, a wide range of low-information values used for our gamma distributions did not change any of our posterior inferences. Therefore, our estimates of odds ratios and their associated 95% credible intervals (which are the Bayesian equivalent of standard confidence intervals) were not unduly affected by our choice of prior distribution. Inferences were calculated with the Gibbs sampler programmed in WinBUGS software (version 1.4, MRC Biostatistics Unit, Cambridge, UK)"

A meta-analysis on implantable defibrillators by Lam and Owen (88) also gives an example of the method, both for choice of priors:

> "To ensure that overall effects were dominated by data from the trials and not influenced by choice of initial distribution we used low information (uninformative) prior distributions—that is, we used vague normal (mean 0, variance 10,000) and uniform (0-2) prior distributions for

means and standard deviations, respectively. We examined the impact of different choices of prior distribution in sensitivity analyses."

and for diagnosis of convergence of chains:

> "The Bayesian models were implemented using WinBUGS version 1.4.1 (Imperial College and Medical Research Council, 2004). After convergence was achieved from an initial 5,000 (burn-in) simulations, we constructed posterior distributions of the treatment effects from three chains of 50,000 simulations. MATLAB version 7.0 (MathWorks, Natick, MA, 2004) was used to carry out diagnostics and further data analyses."

9.3.6 Case study of inconsistency

A network meta-analysis of the effect of antihypertensives on onset of diabetes illustrates the anticipated management of inconsistency in this kind of study (18).
The summary states "With an initial diuretic as the standard of comparison (eight groups), the degree of incoherence (a measure of how closely the entire network fits together) was small (ω=0·000017, eight degrees of freedom)."
In addition, the trial was able to test inconsistency effectively because of the presence of a number of available paths (including paths derived from direct comparisons) for each indirect comparison (Figure 17).
For example, the efficacy of ACE inhibitors compared with thiazide diuretics can be estimated by at least four main paths:

- the direct path ACE inhibitor versus diuretic (2 trials);
- the path ACE inhibitor versus placebo (3 trials) and placebo versus diuretic (3 trials);
- the path ACE inhibitor versus beta-blocker and beta-blocker versus diuretic;
- the path ACE inhibitor versus calcium channel blocker (CCB) (3 trials) and CCB versus diuretic (2 trials);
- and similarly, the paths ACE inhibitor versus CCB, CCB versus ARB and ARB versus diuretic or ACE inhibitor versus CCB, CCB versus placebo and placebo versus diuretic, etc.

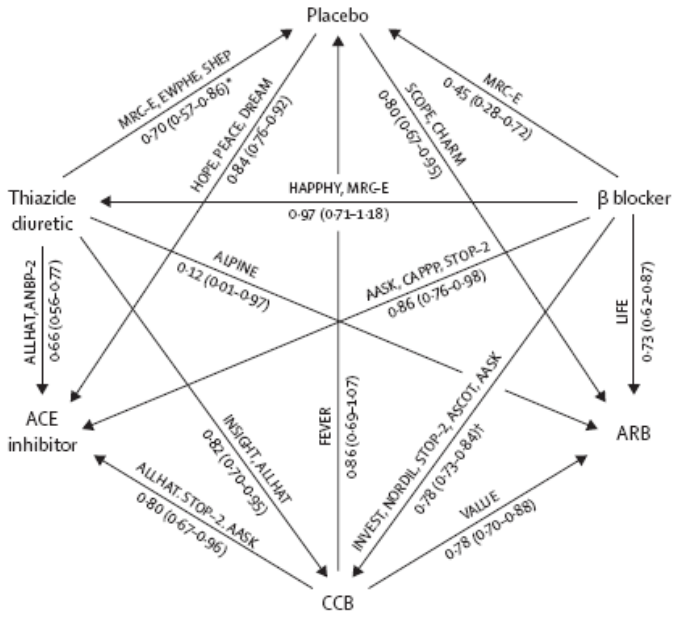


Figure 17 - Network of comparisons of the Elliott and Meyer meta-analysis, 2007

Reprinted from Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. Lancet 2007;369(9557):201-7. Copyright 2007, with permission from Elsevier

9.3.7 Interaction and hypothesis of stability of effects

For thrombolysis during the acute phase of myocardial infarction, the time between onset of symptoms and thrombolysis is a covariable which strongly interacts with the degree of reduction in mortality. The later thrombolysis is started, the smaller the reduction in mortality. A meta-analysis of trials of thrombolysis (89) versus placebo demonstrates this, making it possible to quantify the relationship between time and reduction in mortality.

If a thrombolytic A were evaluated in a placebo-controlled trial in which the mean time from start of symptoms was three hours, its result could not be combined with the result of a placebo-controlled trial of B in which the mean time was 12 hours. In fact it is no longer possible to accept the hypothesis of stability of effects in this configuration.

If the effect of thrombolytics A and B is identical and equal to the mean effect of all thrombolytics measured in the meta-analysis, a reduction of about 28% in mortality will be observed in the trial of A, while the reduction will be 15% in the trial of B, showing the superiority of A over B in the indirect comparison.

10. Conclusion

There are many methods that can be used to carry out indirect comparisons of treatments. These methods make it possible to obtain data on the relative efficacy and safety of the different competing treatments in the complete or partial absence of direct comparison trials (head-to-head trials) of these treatments.

These approaches are therefore potentially of great interest in evaluation as they help to make up for the frequent lack of direct comparisons between treatments. Even when direct comparisons are available, these approaches are still useful as they make it possible to summarise the available data on the relative efficacy and safety of all competing treatments in a given clinical field.

Theoretical analysis of this approach and the methods available, combined with feedback obtained from the first published applications, has made it possible to establish the main outlines of a critical evaluation of studies of this type and to evaluate their reliability in a standard way.

Thus, in view of the technical possibilities, the results of validity studies and the potential benefits of indirect comparisons, may now be envisaged more routinely than before.
The discussion has been initiated, to find out how much indirect comparisons can contribute to health technology assessments and guidelines. Their acceptability needs to be considered in each individual case, taking into account the clinical field and the consequences of an absence of direct comparison trials.

11. Key points

Key points	Section
To make up for the lack of comparative trials of treatments in a large number of clinical fields, the use of indirect comparisons may be considered when there is any question of the ranking of competing treatments in terms of efficacy and/or safety.	Section 1 Section 2.1
This formal approach may make it possible to estimate the relative efficacy of a new treatment compared with existing treatments as soon as the drug has been registered, without having to wait for more conventional comparative data derived from head-to-head trials which become available many years after the product has been placed on the market.	Section 2
The results of direct comparison should not relegate those from indirect comparison to second place; the two estimates should be carefully compared and if possible combined using a mixed approach (mixed treatment comparison)	Section 4.3 Section 6.3
Ideally this approach should be implemented using large network meta-analyses to establish a complete hierarchy in a given clinical field (if necessary, including other resources than drugs alone).	Section 4.4
As there is currently only one large study of the validity of indirect comparisons, further empirical verifications are needed, using modern indirect comparison methods. Research work should be supported to study the questions of reliability of direct comparisons and performance of the various indirect comparison methods proposed.	Section 6.2
Wider diffusion of indirect comparison methods could improve the strategies used to produce a hierarchy of treatments, based on factual data.	Section 1

Annex 1. List of participants

This report was produced by :

- Michel Cucherat, MD, PhD. Evaluation and modeling of therapeutic effects, Biomaths-Health Department, CNRS UMR 5558, Laennec Faculty of Medicine, Claude Bernard Lyon 1 University, Lyon

- Valérie Izard, MD. Department of Medecines Assessment, DEMESP, Haute Autorité de Santé, Saint-Denis

The peer review group consisted of five methodologists :

- Pr Alain Bernard, MD, PhD. Department of Thoracic Surgery, Academic Hospital, Dijon

- Pr Bruno Falissard, MD, PhD. INSERM UMR-S 669, Paris-Sud University, Paris Descartes University & Department of Public Health, Evaluation and Medical Information, Paul-Brousse APHP Academic Hospital, Villejuif

- Patrice Nony, MD, PhD. Department of Clinical Pharmacology, CNRS UMR 5558, Laennec Faculty of Medicine, Cardiovascular Academic Hospital, Lyon

- Pr Eric Vicaut, MD, PhD. Medical Device Evaluation Centre and Clinical Research Unit, Lariboisière Saint-Louis APHP Academic Hospital, Paris

- Mahmoud Zureik, MD, PhD. INSERM U 700, Pathophysiology and Epidemiology of Respiratory Insufficiency, Xavier Bichat Faculty of Medicine, Paris Diderot University, Paris

Annex 2. References

1. Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. Value Health 2008;11(5):956-64.

2. Gartlehner G, Moore CG. Direct versus indirect comparisons: a summary of the evidence. Int J Technol Assess Health Care 2008;24(2):170-7.

3. Malone DC. Using indirect comparisons in pharmacoeconomic studies: time for implementation [editorial]. Clin Ther 2007;29(11):2454-5.

4. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics 2008;26(9):753-67.

5. Pérard L, Hot A, Cucherat M, Simon M, Desmurs H, Coppéré B, *et al.* Essais de non-infériorité dans la maladie thromboembolique veineuse. Une lecture critique est nécessaire! Rev Med Interne 2007;28(11):731-6.

6. Garattini S, Bertelé V. Non-inferiority trials are unethical because they disregard patients' interests. Lancet 2007;370(9602):1875-7.

7. Temple RJ. When are clinical trials of a given agent vs. placebo no longer appropriate or feasible? Control Clin Trials 1997;18(6):613-20.

8. Cour des comptes. La sécurité sociale. La dépense de médicaments 2007. <http://www.ccomptes.fr/fr/CC/documents/RELFSS/07-securite-sociale.pdf> [consulté le 29-10-2008].

9. Brown BM, Hettmansperger TP. Kruskal-Wallis, multiple comparisons and efron dice. Aust N Z J Stat 2002;44(4):427-38.

10. Thangavelu K, Brunner E. Wilcoxon-Mann-Whitney test for stratified samples and Efron's paradox dice. J Stat Plan Inference 2007;137:720-37.

11. Boos DD, Brownie C. Mixture models for continuous data in dose-response studies when some animals are unaffected by treatment. Biometrics 1991;47(4):1489-504.

12. Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, d'Amico R, *et al.* Indirect comparisons of competing interventions. Health Technol Assess 2005;9(26).

13. Julious SA, Wang SJ. How biased are indirect comparisons, particularly when comparisons are made over time in controlled trials? Drug Inf J 2008;42(6):625-33.

14. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 2003;326(7387):472.

15. Soo S, Moayyedi P, Deeks J, Delaney B, Innes M, Forman D. Pharmacological interventions for non-ulcer dyspepsia (review). Cochrane Database of Systematic Reviews 2000;Issue 2.

16. Canadian Coordinating Office for Health Technology Assessment, Trindade E, Menon D. Selective serotonin reuptake inhibitors (SSRIs) for major depression. Part 1. Evaluation of the clinical literature. Ottawa: CCOHTA; 1997.

17. Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. Stat Methods Med Res 2008;17(3):279-301.

18. Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. Lancet 2007;369(9557):201-7.

19. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004;23(20):3105-24.

20. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331(7521):897-900.

21. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002;21(16):2313-24.

22. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. Stat Med 1995;14(24):2685-99.

23. Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996;15(24):2733-49.

24. Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH*, et al.* Health outcomes associated with various antihypertensive therapies used as first-line agents. A network meta-analysis. JAMA 2003;289(19):2534-44.

25. Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-regression. Stat Med 2004;23(11):1663-82.

26. Teramukai S, Matsuyama Y, Mizuno S, Sakamoto J. Individual patient-level and study-level meta-analysis for investigating modifiers of treatment effect. Jpn J Clin Oncol 2004;34(12):717-21.

27. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002;21(11):1559-73.

28. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Stat Med 1999;18(20):2693-708.

29. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med 2002;21(4):589-624.

30. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90 056 participants in 14 randomised trials of statins. Lancet 2005;366(9493):1267-78.

31. Eckert L, Lançon C. Duloxetine compared with fluoxetine and venlafaxine: use of meta-regression analysis for indirect comparisons. BMC Psychiatry 2006;6(30).

32. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. Stat Med 2007;26(6):1237-54.

33. Packer M, Antonopoulos GV, Berlin JA, Chittams J, Konstam MA, Udelson JE. Comparative effects of carvedilol and metoprolol on left ventricular ejection fraction in heart failure: results of a meta-analysis. Am Heart J 2001;141(6):899-907.

34. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Stat Med 2002;21(11):1575-600.

35. Chou R, Carson S, Chan BKS. Gabapentin versus tricyclic antidepressants for diabetic neuropathy and post-herpetic neuralgia: discrepancies between direct and indirect meta-analyses of randomized controlled trials. J Gen Intern Med 2009;24(2):178-88.

36. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JPT, Churchill R*, et al.* Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 2009;373(9665):746-58.

37. Edwards SJ, Clarke MJ, Wordsworth S, Welton NJ. Carbapenems versus other beta-lactams in the

treatment of hospitalised patients with infection: a mixed treatment comparison. Curr Med Res Opin 2009;25(1):251-61.

38. Karthikeyan G, Eikelboom JW, Yusuf S, Hirsh J. An indirect comparison of the efficacy and safety of factor Xa inhibitors with thrombin inhibitors nn preventing venous thromboembolism after hip or knee surgery [abstract]. Circulation 2008;118(Suppl 2):S1051.

39. Lee YH, Woo JH, Rho YH, Choi SJ, Ji JD, Song GG. Meta-analysis of the combination of TNF inhibitors plus MTX compared to MTX monotherapy, and the adjusted indirect comparison of TNF inhibitors in patients suffering from active rheumatoid arthritis. Rheumatol Int 2008;28(6):553-9.

40. Mauri D, Polyzos NP, Salanti G, Pavlidis N, Ioannidis JPA. Multiple-treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer. J Natl Cancer Inst 2008;100(24):1780-91.

41. Biondi-Zoccai GGL, Lotrionte M, Abbate A, Valgimigli M, Testa L, Burzotta F*, et al.* Direct and indirect comparison meta-analysis demonstrates the superiority of sirolimus- versus paclitaxel-eluting stents across 5854 patients. Int J Cardiol 2007;114(1):104-5.

42. Boonen S, Lips P, Bouillon R, Bischoff-Ferrari HA, Vanderschueren D, Haentjens P. Need for additional calcium to reduce the risk of hip fracture with vitamin D supplementation: evidence from a comparative metaanalysis of randomized controlled trials. J Clin Endocrinol Metab 2007;92(4):1415-23.

43. Purkayastha S, Athanasiou T, Tekkis PP, Constantinides V, Teare J, Darzi AW. Magnetic resonance colonography *vs* computed tomography colonography for the diagnosis of colorectal cancer: an indirect comparison. Colorectal Dis 2007;9(2):100-11.

44. Kamphuisen PW, Agnelli G. What is the optimal pharmacological prophylaxis for the prevention of deep-vein thrombosis and pulmonary embolism in patients with acute ischemic stroke? Thromb Res 2007;119(3):265-74.

45. Lee YH, Ji JD, Song GG. Adjusted indirect comparison of celecoxib versus rofecoxib on cardiovascular risk. Rheumatol Int 2006;27:477-82.

46. Griffin S, Bojke L, Main C, Palmer S. Incorporating direct and indirect evidence using bayesian methods: an applied case study in ovarian cancer. Value Health 2006;9(2):123-31.

47. Zhou Z, Rahme E, Pilote L. Are statins created equal? Evidence from randomized trials of pravastatin, simvastatin, and atorvastatin for cardiovascular disease prevention. Am Heart J 2006;151(2):273-81.

48. Stettler C, Allemann S, Egger M, Windecker S, Meier B, Diem P. Efficacy of drug eluting stents in patients with and without diabetes mellitus: indirect comparison of controlled trials. Heart 2006;92(5):650-7.

49. Berner MM, Kriston L, Harms A. Efficacy of PDE-5-inhibitors for erectile dysfunction. A comparative meta-analysis of fixed-dose regimen randomized controlled trials administering the International Index of Erectile Function in broad-spectrum populations. Int J Impot Res 2006;18(3):229-35.

50. Batterham MJ. Investigating heterogeneity in studies of resting energy expenditure in persons with HIV/AIDS: a meta-analysis. Am J Clin Nutr 2005;81(3):702-13.

51. Biondi-Zoccai GGL, Agostoni P, Abbate A, Testa L, Burzotta F, Lotrionte M*, et al.* Adjusted indirect comparison of intracoronary drug-eluting stents: evidence from a metaanalysis of randomized bare-metal-stent-controlled trials. Int J Cardiol 2005;100(1):119-23.

52. Berry C, Norrie J, McMurray JJV. Ximelagatran compared with warfarin for the prevention of systemic embolism and stroke. An imputed placebo

analysis. Cardiovasc Drugs Ther 2005;19(2):149-51.

53. Moore RA, Derry S, McQuay HJ. Indirect comparison of interventions using published randomised trials: systematic review of PDE-5 inhibitors for erectile dysfunction. BMC Urol 2005;5(18).

54. Babapulle MN, Joseph L, Bélisle P, Brophy JM, Eisenberg MJ. A hierarchical Bayesian meta-analysis of randomised clinical trials of drug-eluting stents. Lancet 2004;364(9434):583-91.

55. Yazdanpanah Y, Sissoko D, Egger M, Mouton Y, Zwahlen M, Chêne G. Clinical efficacy of antiretroviral combination therapy based on protease inhibitors or non-nucleoside analogue reverse transcriptase inhibitors: indirect comparison of controlled trials. BMJ 2004;328(7434):249.

56. Jones L, Griffin S, Palmer S, Main C, Orton V, Sculpher M*, et al.* Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation. Health Technol Assess 2004;8(38).

57. Hind D, Calvert N, McWilliams R, Davidson A, Paisley S, Beverley C*, et al.* Ultrasonic locating devices for central venous cannulation: meta-analysis. BMJ 2003;327(7411):361.

58. Lim E, Ali Z, Ali A, Routledge T, Edmonds L, Altman DG*, et al.* Indirect comparison meta-analysis of aspirin therapy after coronary surgery. BMJ 2003;327(7427):1309.

59. Coomarasamy A, Knox EM, Gee H, Song F, Khan KS. Effectiveness of nifedipine *versus* atosiban for tocolysis in preterm labour: a meta-analysis with an indirect comparison of randomised trials. BJOG 2003;110(12):1045-9.

60. Bhandari M, Guyatt GH, Swiontkowski MF, Schemitsch EH. Treatment of open fractures of the shaft of the tibia. A systematic overview and meta-analysis. J Bone Joint Surg [Br] 2001;83B(1):62-8.

61. Agency for Healthcare Research and Quality. Manifestations and management of chronic insomnia in adults. Evidence report/Technology assessment n° 125. Rockville: AHRQ; 2005.

62. BlueCross BlueShield Association. Metal-on-metal total hip resurfacing. Technol Eval Cent Asses Program Exec Summ 2007;22(3):1-44.

63. BlueCross BlueShield Association. Off-label uses of bevacizumab: renal cell carcinoma and other miscellaneous non-colorectal cancer indications. Technol Eval Cent Asses Program Exec Summ 2006;21(9):1-26.

64. Costa J, Espírito-Santo C, Borges A, Ferreira JJ, Coelho M, Moore P*, et al.* Botulinum toxin type A therapy for cervical dystonia (Review). Cochrane Database of Systematic Reviews 2005;Issue 1.

65. Macfadyen CA, Acuin JM, Gamble C. Topical antibiotics without steroids for chronically discharging ears with underlying eardrum perforations (Review). Cochrane Database of Systematic Reviews 2005;Issue 4.

66. National Institute for Health and Clinical Excellence. Docetaxel for the adjuvant treatment of early node-positive breast cancer. NICE technology appraisal guidance 109. London: NICE; 2006.

67. National Institute for Health and Clinical Excellence. Pemetrexed for the treatment of non-small-cell lung cancer. NICE technology appraisal guidance 124. London: NICE; 2007.

68. National Institute for Health and Clinical Excellence. Adalimumab for the treatment of psoriatic arthritis. NICE technology appraisal guidance 125. London: NICE; 2007.

69. National Institute for Health and Clinical Excellence. Rituximab for the treatment of

rheumatoid arthritis. NICE technology appraisal guidance 126. London: NICE; 2007.

70. Agency for Healthcare Research and Quality. Comparative effectiveness and safety of oral diabetes medications for adults with type 2 diabetes. Rockville: AHRQ; 2007.

71. National Institute for Health and Clinical Excellence. Infliximab for the treatment of adults with psoriasis. NICE technology appraisal guidance 134. London: NICE; 2008.

72. National Institute for Health and Clinical Excellence. Varenicline for smoking cessation. NICE technology appraisal guidance 123. London: NICE; 2007.

73. Zhang WY, Li Wan Po A. Analgesic efficacy of paracetamol and its combination with codeine and caffeine in surgical pain: a meta-analysis. J Clin Pharm Ther 1996;21(4):261-82.

74. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997;50(6):683-91.

75. Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy illustrated by antimicrobial prophylaxis in colorectal surgery. Control Clin Trials 2000;21(5):488-97.

76. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003;326(7400):1167-70.

77. Salonen R. Drug comparisons: why are they so difficult? Cephalalgia 2000;20(Suppl 2):25-32.

78. Fuseau E, Petricoul O, Sabin A, Pereira A, O'Quinn S, Thein S*, et al.* Effect of encapsulation on absorption of sumatriptan tablets: data from healthy volunteers and patients during a migraine. Clin Ther 2001;23(2):242-51.

79. Pascual J, Mateos V, Roig C, Sanchez-del-Rio M, Jiménez D. Marketed oral triptans in the acute treatment of migraine: a systematic review on efficacy and tolerability. Headache 2007;47(8):1152-68.

80. Jones RW, Soininen H, Hager K, Aarsland D, Passmore P, Murthy A*, et al.* A multinational, randomised, 12-week study comparing the effects of donepezil and galantamine in patients with mild to moderate Alzheimer's disease. Int J Geriatr Psychiatry 2004;19(1):58-67.

81. Wilcock G, Howe I, Coles H, Lilienfeld S, Truyen L, Zhu Y*, et al.* A long-term comparison of galantamine and donepezil in the treatment of Alzheimer's disease. Drugs Aging 2003;20(10):777-89.

82. Song F, Harvey I, Lilford R. Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. J Clin Epidemiol 2008;61(5):455-63.

83. National Institute for Health and Clinical Excellence. Specification for manufacturer/sponsor submission of evidence. Single technology appraisal. London: NICE; 2006.

84. National Institute for Clinical Excellence. Guide to the methods of technology appraisal. London: NICE; 2004.

85. Wells GA, Sultan SA, Chen L, Khan M, Coyle D. Indirect evidence: indirect treatment comparisons in meta-analysis. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009.

86. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ 2009;338:b1147.

87. Van der Valk R, Webers CAB, Schouten JSAG, Zeegers MP, Hendrikse F, Prins MH. Intraocular

pressure-lowering effects of all commonly used glaucoma drugs. A meta-analysis of randomized clinical trials. Ophthalmology 2005;112(7):1177-85.

88. Lam SKH, Owen A. Combined resynchronisation and implantable defibrillator therapy in left ventricular dysfunction: Bayesian network meta-analysis of randomised controlled trials. BMJ 2007;335(7626):925.

89. Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. Lancet 1994;343(8893):311-22.

Annex 3. Search for guidelines, health technology assessments and methodological procedures on the Internet; list of websites consulted

Agence Française de Sécurité Sanitaire des Produits de Santé - AFSSAPS
Lemanissier Medical library
Bibliothèque InterUniversitaire de Médecine - BIUM
Catalogue et Index des Sites Médicaux Francophones - CISMeF
Comité d'Evaluation et de Diffusion des Innovations Technologiques - CEDIT
INSERM Collective Expertise
Fédération Nationale des Centres de Lutte Contre le Cancer - FNCLCC
Société Française de Médecine Générale - SFMG
Adelaide Health Technology Assessment - AHTA
Agence d'Evaluation des Technologies et des Modes d'Intervention en Santé - AETMIS
Agency for Healthcare Research and Quality - AHRQ
Alberta Heritage Foundation for Medical Research - AHFMR
Australian Safety and Efficacy Register of New Interventional Procedures - Surgical
Blue Cross Blue Shield Association - BCBS - Technology Evaluation Center
Canadian Agency for Drugs and Technologies in Health - CADTH
Cancer Care Ontario
Centers for Disease Control and Prevention - CDC
Centre fédéral d'expertise des soins de santé - KCE
Centre for Clinical Effectiveness - CCE
Centre for Reviews and Dissemination databases
Clinical Knowledge Summaries
Canadian Medical Association Infobase
Cochrane Library
European Medicines Agency - EMA
Food and Drug Administration - FDA
Guideline Advisory Committee - GAC
Guidelines and Protocols Advisory Committee - GPAC
Guidelines International Network - GIN
National Library for Health Guidelines Finder
Health Services Technology Assessment Text - HSTAT
Horizon Scanning
Institute for Clinical Evaluative Sciences - ICES
Institute for clinical systems improvement.
Institute for Health Economics Alberta - IHE
Intute Health & Life Sciences - INTUTE
Medical Services Advisory Committee - MSAC
Minnesota Department of Health - Health Technology Advisory Committee - HTAC
National Comprehensive Cancer Network
National Coordinating Centre for Health Technology Assessment - NCCHTA
National Guideline Clearinghouse - NGC
National Health Services Scotland
National Institute for Health and Clinical Excellence - NICE
National Institutes of Health - NIH
New Zealand Health Technology Assessment - NZHTA
Ontario Health Technology Advisory Committee - OHTAC
Scottish Intercollegiate Guidelines Network - SIGN
Singapore Ministry of Health
U.S. Preventive Services Task Force - USPSTF
Veterans Affairs Technology Assessment Program
West Midlands Health Technology Assessment Collaboration - WMHTA

Index

A

a priori, 51
a priori non informatif, 28
adjusted indirect comparison, 18
ampleur d'effet, 35
approche mixte, 22
approche naïve, 16
assay sensitivity, 44

C

chaînes de Markov, 50
Comparaison des différentes méthodes, 35
comparaison directe, 9
comparaison face à face, 9
comparaison indirecte, 9
comparaison indirecte ajustée, 52
comparaison naïve, 51, 52
comparaisons directes, 51
concordance comparaison directe - comparaison indirecte, 51
convergence, 50
Cour des Comptes, 12
critères de jugement continus, 35

D

diagnostic de convergence, 50
discordance, 41
discrepancy, 41
distribution a posteriori, 28
donepezil, 45
double aveugle, 43
double placebo, 44

E

effect size, 35
EMEA, 46
essai de non infériorité, 12
essai versus traitement actif, 43
essais avec control actif, 9
essais de non infériorités, 13
estimation indirecte ajustée, 18

F

face à face, 43
FDA, 46
fibrinolyse, 54
fibrinolytique, 55

G

galantamine, 45
Gibbs sampling, 50

H

head to head comparison, 9
hypothèse de stabilité des effets, 19, 49, 50, 54

I

ICH E9, 13
infarctus du myocarde, 11, 54

L

lois de financement de la sécurité sociale, 12

M

maladie d'Alzheimer, 11, 45
MCMC, 50, 53
migraine, 11
modèle de Cox, 13
Monte Carlo, 50
Monte Carlo par chaînes de Markov, 53

N

NICE, 46
non transitivité des tests statistiques, 13

O

odds ratio, 36

R

risque relatif, 36
risque relatif., 9

S

sécurité sociale, 12
stabilité des effets, 19, 49, 50, 54

T

test de Wilcoxon, 13
test du logrank, 13
théorème de Bayes, 28
traitements concurrents, 11

V

validité, 41

W

weighted mean difference, 35

HAUTE AUTORITÉ DE SANTÉ

All HAS publications can be downloaded from www.has-sante.fr